(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number
WO 02/08392 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/EP01/08182

(22) International Filing Date: 16 July 2001 (16.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/220,807	25 July 2000 (25.07.2000)	US
60/280,109	2 April 2001 (02.04.2001)	US

(71) Applicant *(for all designated States except US)*: **BAYER AKTIENGESELLSCHAFT** [DE/DE]; 51368 Leverkusen (DE).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **XIAO, Yonghong** [US/US]; 78 Dana Street #1, Cambridge, MA 02138 (US).

(74) Common Representative: **BAYER AKTIENGESELLSCHAFT**; 51368 Leverkusen (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/08392 A2

(54) Title: REGULATION OF HUMAN MATRIPTASE-LIKE SERINE PROTEASE

(57) Abstract: Reagents which regulate human matriptase-like serine protease activity and reagents which bind to human matriptase-like serine protease gene products can be used to regulate extracellular matrix degradation. Such regulation is particularly useful for treating cancer, including metastasis of malignant cells, chronic obstructive pulmonary disease, disorders of the peripheral or central nervous system, and cardiovascular disorders.

REGULATION OF HUMAN MATRIPTASE-LIKE SERINE PROTEASE

TECHNICAL FIELD OF THE INVENTION

The invention relates to the area of regulation of extracellular matrix degradation. More particularly, the invention relates to the regulation of human matriptase-like serine protease activity to increase or decrease extracellular matrix degradation.

BACKGROUND OF THE INVENTION

Metastasizing cancer cells invade the extracellular matrix using plasma membrane protrusions that contact and dissolve the matrix with proteases. Agents which inhibit such protease activity can be used to suppress metastases. Proteases also are expressed during development, when degradation of the extracellular matrix is desired. In cases where appropriate extracellular matrix degradation does not occur, supplying a molecule with a protease activity can provide the necessary enzymatic activity. Thus, there is a need in the art for identifying new proteases and methods of regulating extracellular matrix degradation.

SUMMARY OF THE INVENTION

It is an object of the invention to provide reagents and methods of regulating a human matriptase-like serine protease. This and other objects of the invention are provided by one or more of the embodiments described below.

One embodiment of the invention is a matriptase-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2;

the amino acid sequence shown in SEQ ID NO: 2;

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 18; and

the amino acid sequence shown in SEQ ID NO: 18.

Yet another embodiment of the invention is a method of screening for agents which decrease extracellular matrix degradation. A test compound is contacted with a matriptase-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2; and

the amino acid sequence shown in SEQ ID NO: 2;

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 18; and

the amino acid sequence shown in SEQ ID NO: 18.

Binding between the test compound and the matriptase-like serine protease polypeptide is detected. A test compound which binds to the matriptase-like serine protease polypeptide is thereby identified as a potential agent for decreasing extracellular matrix degradation. The agent can work by decreasing the activity of the matriptase-like serine protease.

Another embodiment of the invention is a method of screening for agents which decrease extracellular matrix degradation. A test compound is contacted with a polynu-

cleotide encoding a matriptase-like serine protease polypeptide, wherein the polynucleotide comprises a nucleotide sequence selected from the group consisting of:

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 1;

the nucleotide sequence shown in SEQ ID NO: 1;

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 17; and

the nucleotide sequence shown in SEQ ID NO: 17.

Binding of the test compound to the polynucleotide is detected. A test compound which binds to the polynucleotide is identified as a potential agent for decreasing extracellular matrix degradation. The agent can work by decreasing the amount of the matriptase-like serine protease through interacting with the matriptase-like serine protease mRNA.

Another embodiment of the invention is a method of screening for agents which regulate extracellular matrix degradation. A test compound is contacted with a matriptase-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2; and

the amino acid sequence shown in SEQ ID NO: 2;

amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 18; and

the amino acid sequence shown in SEQ ID NO: 18.

A matriptase-like serine protease activity of the polypeptide is detected. A test compound which increases matriptase-like serine protease activity of the polypeptide relative to matriptase-like serine protease activity in the absence of the test compound is thereby identified as a potential agent for increasing extracellular matrix degradation. A test compound which decreases matriptase-like serine protease activity of the polypeptide relative to matriptase-like serine protease activity in the absence of the test compound is thereby identified as a potential agent for decreasing extracellular matrix degradation.

Even another embodiment of the invention is a method of screening for agents which decrease extracellular matrix degradation. A test compound is contacted with a matriptase-like serine protease product of a polynucleotide which comprises a nucleotide sequence selected from the group consisting of:

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 1;

the nucleotide sequence shown in SEQ ID NO: 1;

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 17; and

the nucleotide sequence shown in SEQ ID NO: 17.

Binding of the test compound to the matriptase-like serine protease product is detected. A test compound which binds to the matriptase-like serine protease product is thereby identified as a potential agent for decreasing extracellular matrix degradation.

Still another embodiment of the invention is a method of reducing extracellular matrix degradation. A cell is contacted with a reagent which specifically binds to a polynucleotide encoding a matriptase-like serine protease polypeptide or the product encoded by the polynucleotide, wherein the polynucleotide comprises a nucleotide sequence selected from the group consisting of:

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 1;

the nucleotide sequence shown in SEQ ID NO: 1;

nucleotide sequences which are at least about 50% identical to the nucleotide sequence shown in SEQ ID NO: 17; and

the nucleotide sequence shown in SEQ ID NO: 17.

Matriptase-like serine protease activity in the cell is thereby decreased.

The invention thus provides reagents and methods for regulating extracellular matrix degradation which can be used *inter alia*, to suppress metastatic activity of malignant cells and to enhance extracellular matrix degradation during development.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the DNA-sequence encoding a matriptase-like serine protease polypeptide (SEQ ID NO:1)

Fig. 2 shows the amino acid sequence deduced from the DNA-sequence of Fig. 1 (SEQ ID NO:2)

Fig. 3 shows the DNA-sequence of a protein identified by EMBL Accession No. AF118224 (SEQ ID NO:3)

Fig. 4 shows the amino acid sequence of pfam/hmm/trypsin (SEQ ID NO:15)

Fig. 5 shows the amino acid sequence of pfam/hmm/ldl_recept_a-low-density lipoprotein receptor domain (SEQ ID NO:16)

Fig.6 shows the DNA-sequence encoding a matriptase-like serine protease polypeptide (SEQ ID NO:17)

Fig. 7 shows the BLASTP alignment of SEQ ID NO:2 with SEQ ID NO:3.

Fig. 8 shows the Prosite search results.

Fig. 9 shows the BLOCKS search results.

Fig. 10 shows the HMMPFAM alignment of SEQ ID NO:2 with pfam|hmm|trypsin (SEQ ID NO:15).

Fig. 11 shows the HMMPFAM alignment of SEQ ID NO:2 with pfam|hmm|ldl_recept_a low-density lipoprotein receptor domain (SEQ ID NO:16).

Fig. 12 shows the relative expression of human matriptase-like serine protease in respiratory cells and tissues.

Fig. 13 shows the relative expression of human matriptase-like serine protease I in various human tissues and the neutrophil-like cell line HL60.

Fig. 14 shows the BLASTP-alignment of 151_ext2_protein against trembl|AF118224|AF118224_1

Fig. 15 shows that the BLOCKS search results are the same as the previous version of 151

Fig. 16 shows the HMMPFAM-alignment of 151_ext2_protein against pfam/hmm/trypsin

Fig. 17 shows the HMMPFAM-alignment of 151_ext2_protein against pfam/hmm/ldl_recept_a.

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to an isolated polynucleotide encoding a matriptase-like serine protease polypeptide and being selected from the group consisting of:

a) a polynucleotide encoding a matriptase-like serine protease polypeptide comprising an amino acid sequence selected from the group consisting of:

amino acid sequences which are at least about 50% identical to
the amino acid sequence shown in SEQ ID NO: 2;
the amino acid sequence shown in SEQ ID NO: 2;
amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 18; and
the amino acid sequence shown in SEQ ID NO: 18.

b) a polynucleotide comprising the sequence of SEQ ID NOS: 1 or 17;

c) a polynucleotide which hybridizes under stringent conditions to a polynucleotide specified in (a) and (b);

d) a polynucleotide the sequence of which deviates from the polynucleotide sequences specified in (a) to (c) due to the degeneration of the genetic code; and

e) a polynucleotide which represents a fragment, derivative or allelic variation of a polynucleotide sequence specified in (a) to (d).

Furthermore, it has been discovered by the present applicant that regulators of a matriptase-like serine protease, particularly human matriptase-like serine protease, can be used to regulate degradation of the extracellular matrix. Human matriptase-like serine protease as shown in SEQ ID NO:2 is 44% identical over 294 amino acids to the protein identified by EMBL Accession No. AF118224 (SEQ ID NO:3) and annotated as a *Homo sapiens* matriptase (FIG. 1). Human matriptase-like serine protease contains domains typical of a trypsin family serine protease (Figs. 2-5).

A coding sequence for SEQ ID NO:2 is shown in SEQ ID NO:1. This coding sequence is found within genomic clones identified with GenBank Accession Nos. AC010617, AC011542, and AC011522. An EST identified with EMBL Accession No. AI698086 (463bp) is contained within SEQ ID NO:1, indicating that human matriptase-like serine protease is expressed.

Human matriptase-like serine protease is expected to be useful for the same purposes as previously identified matriptase serine protease (*see* Lin *et al.*, *J. Biol. Chem. 274*, 18231-36, 1999, and Lin *et al.*, *J. Biol. Chem. 274*, 18237-42, 1999). Human ma-

triptase-like serine protease is expected to be especially useful for treating cancer, chronic obstructive pulmonary disease (COPD), disorders of the peripheral or central nervous system, and cardiovascular disorders.

Polypeptides

Matriptase-like serine protease polypeptides according to the invention comprise an amino acid sequence as shown in SEQ ID NO:2 and 18, a portion of SEQ ID NO:2 or 18 comprising at least 6, 10, 15, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, or 340 contiguous amino acids, or a biologically active variant of the amino acid sequence shown in SEQ ID NO:2 or 18, as defined below. A matriptase-like serine protease polypeptide of the invention therefore can be a portion of a matriptase-like serine protease molecule, a full-length matriptase-like serine protease molecule, or a fusion protein comprising all or a portion of a matriptase-like serine protease molecule.

Biologically Active Variants

Matriptase-like serine protease variants which are biologically active, *i.e.*, retain a matriptase-like serine protease activity, also are matriptase-like serine protease polypeptides. Preferably, naturally or non-naturally occurring matriptase-like serine protease variants have amino acid sequences which are at least about 50, 55, 60, 65, 70, preferably about 75, 90, 96, or 98% identical to an amino acid sequence shown in SEQ ID NO:2 or 18. Percent identity between a putative matriptase-like serine protease variant and an amino acid sequence of SEQ ID NO:2 and 18 is determined with the Needleman/Wunsch algorithm (Needleman and Wunsch, J.Mol. Biol. 48; 443-453, 1970) using a Blosum62 matrix with a gap creation penalty of 8 and a gap extension penalty of 2 (S. Henikoff and J.G. Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-10919, 1992).

Variations in percent identity can be due, for example, to amino acid substitutions, insertions, or deletions. Amino acid substitutions are defined as one for one amino acid replacements. They are conservative in nature when the substituted amino acid has similar structural and/or chemical properties. Examples of conservative replacements are substitution of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine.

Amino acid insertions or deletions are changes to or within an amino acid sequence. They typically fall in the range of about 1 to 5 amino acids. Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, such as DNASTAR software. Whether an amino acid change results in a biologically active matriptase-like serine protease polypeptide can readily be determined by assaying for matriptase-like serine protease activity, as is known in the art and described, for example, in Lin *et al.*, *J. Biol. Chem. 274*, 18231-36, 1999.

Fusion Proteins

Fusion proteins are useful for generating antibodies against matriptase-like serine protease amino acid sequences and for use in various assay systems. For example, fusion proteins can be used to identify proteins which interact with portions of a matriptase-like serine protease polypeptide, including its active site and fibronectin domains. Methods such as protein affinity chromatography or library-based assays for protein-protein interactions, such as the yeast two-hybrid or phage display systems, can be used for this purpose. Such methods are well known in the art and also can be used as drug screens.

A matriptase-like serine protease fusion protein comprises two protein segments fused together by means of a peptide bond. Contiguous amino acids for use in a fusion protein can be selected from the amino acid sequence shown in SEQ ID NO:2 or 18or from a biologically active variants of those sequences, such as those described

above. For example, the first protein segment can comprise at least 6, 10, 15, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, or 340 or more contiguous amino acids of SEQ ID NO:2 or a biologically active variant. Preferably, a fusion protein comprises the active site of the protease and/or one or both of the fibronectin domains. The first protein segment also can comprise full-length matriptase-like serine protease.

The second protein segment can be a full-length protein or a protein fragment or polypeptide. Proteins commonly used in fusion protein construction include β-galactosidase, β-glucuronidase, green fluorescent protein (GFP), autofluorescent proteins, including blue fluorescent protein (BFP), glutathione-S-transferase (GST), luciferase, horseradish peroxidase (HRP), and chloramphenicol acetyltransferase (CAT). Additionally, epitope tags are used in fusion protein constructions, including histidine (His) tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Other fusion constructions can include maltose binding protein (MBP), S-tag, Lex a DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. A fusion protein also can be engineered to contain a cleavage site located between the matriptase-like serine protease polypeptide-encoding sequence and the heterologous protein sequence, so that the matriptase-like serine protease polypeptide can be cleaved and purified away from the heterologous moiety.

A fusion protein can be synthesized chemically, as is known in the art. Preferably, a fusion protein is produced by covalently linking two protein segments or by standard procedures in the art of molecular biology. Recombinant DNA methods can be used to prepare fusion proteins, for example, by making a DNA construct which comprises matriptase-like serine protease coding sequences disclosed herein in proper reading frame with nucleotides encoding the second protein segment and expressing the DNA construct in a host cell, as is known in the art. Many kits for constructing fusion proteins are available from companies such as Promega Corporation (Madison, WI), Stratagene (La Jolla, CA), CLONTECH (Mountain View, CA), Santa Cruz

Biotechnology (Santa Cruz, CA), MBL International Corporation (MIC; Watertown, MA), and Quantum Biotechnologies (Montreal, Canada; 1-888-DNA-KITS).

Identification of Species Homologs

Species homologs of human matriptase-like serine protease can be obtained using matriptase-like serine protease polynucleotides (described below) to make suitable probes or primers to screening cDNA expression libraries from other species, such as mice, monkeys, or yeast, identifying cDNAs which encode homologs of matriptase-like serine protease, and expressing the cDNAs as is known in the art.

Polynucleotides

A matriptase-like serine protease polynucleotide can be single- or double-stranded and comprises a coding sequence or the complement of a coding sequence for a matriptase-like serine protease polypeptide. A partial coding sequence of a matriptase-like serine protease polynucleotide is shown in SEQ ID NO:1.

Degenerate nucleotide sequences encoding human matriptase-like serine protease polypeptides, as well as homologous nucleotide sequences which are at least about 50, 55, 60, 65, 70, preferably about 75, 90, 96, or 98% identical to the matriptase-like serine protease coding sequences nucleotide sequence shown in SEQ ID NO:1 or 17 also are matriptase-like serine protease polynucleotides. Percent sequence identity between the sequences of two polynucleotides is determined using computer programs such as ALIGN which employ the FASTA algorithm, using an affine gap search with a gap open penalty of -12 and a gap extension penalty of -2. Complementary DNA (cDNA) molecules, species homologs, and variants of matriptase-like serine protease polynucleotides which encode biologically active matriptase-like serine protease polypeptides also are matriptase-like serine protease polynucleotides.

Identification of Variants and Homologs

Variants and homologs of the matriptase-like serine protease polynucleotides disclosed above also are matriptase-like serine protease polynucleotides. Typically, homologous matriptase-like serine protease polynucleotide sequences can be identified by hybridization of candidate polynucleotides to known matriptase-like serine protease polynucleotides under stringent conditions, as is known in the art. For example, using the following wash conditions--2X SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7.0), 0.1% SDS, room temperature twice, 30 minutes each; then 2X SSC, 0.1% SDS, 50 °C once, 30 minutes; then 2X SSC, room temperature twice, 10 minutes each--homologous sequences can be identified which contain at most about 25-30% basepair mismatches. More preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches.

Species homologs of the matriptase-like serine protease polynucleotides disclosed herein can be identified by making suitable probes or primers and screening cDNA expression libraries from other species, such as mice, monkeys, or yeast. Human variants of matriptase-like serine protease polynucleotides can be identified, for example, by screening human cDNA expression libraries. It is well known that the T_m of a double-stranded DNA decreases by 1-1.5 °C with every 1% decrease in homology (Bonner *et al.*, *J. Mol. Biol. 81*, 123 (1973). Variants of human matriptase-like serine protease polynucleotides or matriptase-like serine protease polynucleotides of other species can therefore be identified, for example, by hybridizing a putative homologous matriptase-like serine protease polynucleotide with a polynucleotide having a nucleotide sequence of SEQ ID NO:1 or 17 or an ephrin-like serine protease coding sequence of SEQ ID NO: 3 to form a test hybrid. The melting temperature of the test hybrid is compared with the melting temperature of a hybrid comprising matriptase-like serine protease polynucleotides having perfectly complementary nucleotide sequences, and the number or percent of basepair mismatches within the test hybrid is calculated.

Nucleotide sequences which hybridize to matriptase-like serine protease polynucleotides or their complements following stringent hybridization and/or wash conditions are also matriptase-like serine protease polynucleotides. Stringent wash conditions are well known and understood in the art and are disclosed, for example, in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed., 1989, at pages 9.50-9.51.

Typically, for stringent hybridization conditions a combination of temperature and salt concentration should be chosen that is approximately 12-20 °C below the calculated T_m of the hybrid under study. The T_m of a hybrid between a matriptase-like serine protease polynucleotide having a coding sequence disclosed herein and a polynucleotide sequence which is at least about 50, preferably about 75, 90, 96, or 98% identical to that nucleotide sequence can be calculated, for example, using the equation of Bolton and McCarthy, *Proc. Natl. Acad. Sci. U.S.A. 48*, 1390 (1962):

$$T_m = 81.5\ ^{\circ}C - 16.6(\log_{10}[Na^+]) + 0.41(\%G + C) - 0.63(\%\text{formamide}) - 600/l),$$

where l = the length of the hybrid in basepairs.

Stringent wash conditions include, for example, 4X SSC at 65 °C, or 50% formamide, 4X SSC at 42 °C, or 0.5X SSC, 0.1% SDS at 65 °C. Highly stringent wash conditions include, for example, 0.2X SSC at 65 °C.

Preparation of Polynucleotides

A naturally occurring matriptase-like serine protease polynucleotide can be isolated free of other cellular components such as membrane components, proteins, and lipids. Polynucleotides can be made by a cell and isolated using standard nucleic acid purification techniques, synthesized using an amplification technique, such as the polymerase chain reaction (PCR), or synthesized using an automatic synthesizer. Methods for isolating polynucleotides are routine and are known in the art. Any such

technique for obtaining a polynucleotide can be used to obtain isolated matriptase-like serine protease polynucleotides. For example, restriction enzymes and probes can be used to isolate polynucleotide fragments which comprise matriptase-like serine protease nucleotide sequences. Isolated polynucleotides are in preparations which are free or at least 70, 80, or 90% free of other molecules.

Matriptase-like serine protease cDNA molecules can be made with standard molecular biology techniques, using matriptase-like serine protease mRNA as a template. Matriptase-like serine protease cDNA molecules can thereafter be replicated using molecular biology techniques known in the art and disclosed in manuals such as Sambrook *et al.* (1989). An amplification technique, such as PCR, can be used to obtain additional copies of matriptase-like serine protease polynucleotides, using either human genomic DNA or cDNA as a template.

Alternatively, synthetic chemistry techniques can be used to synthesize matriptase-like serine protease polynucleotides. The degeneracy of the genetic code allows alternate nucleotide sequences to be synthesized which will encode a matriptase-like serine protease polypeptide having, for example, the amino acid sequence shown in SEQ ID NO:2 or 18 or a biologically active variant of that sequence.

Obtaining Full-Length Polynucleotides

The partial sequence of SEQ ID NO:1 or its complement can be used to identify the corresponding full length gene from which they were derived. The partial sequences can be nick-translated or end-labeled with ^{32}P using polynucleotide kinase using labeling methods known to those with skill in the art (BASIC METHODS IN MOLECULAR BIOLOGY, Davis *et al.*, eds., Elsevier Press, N.Y., 1986). A lambda library prepared from human tissue can be directly screened with the labeled sequences of interest or the library can be converted en masse to pBluescript (Stratagene Cloning Systems, La Jolla, Calif. 92037) to facilitate bacterial colony screening (see Sambrook *et al.*, 1989, pg. 1.20).

Both methods are well known in the art. Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured, and the DNA is fixed to the filters. The filters are hybridized with the labeled probe using hybridization conditions described by Davis *et al.*, 1986. The partial sequences, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive colony or plaque. The colonies or plaques are selected and expanded, and the DNA is isolated from the colonies for further analysis and sequencing.

Positive cDNA clones are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the partial sequence and the other primer from the vector. Clones with a larger vector-insert PCR product than the original partial sequence are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size determined from Northern blot Analysis.

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined, for example after exonuclease III digestion (McCombie *et al.*, *Methods 3*, 33-40, 1991). A series of deletion clones are generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

Various PCR-based methods can be used to extend the nucleic acid sequences encoding the disclosed portions of human matriptase-like serine protease to detect upstream sequences such as promoters and regulatory elements. For example, restriction-site PCR uses universal primers to retrieve unknown sequence adjacent to a

known locus (Sarkar, *PCR Methods Applic. 2*, 318-322, 1993). Genomic DNA is first amplified in the presence of a primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR also can be used to amplify or extend sequences using divergent primers based on a known region (Triglia *et al.*, *Nucleic Acids Res. 16*, 8186, 1988). Primers can be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, Minn.), to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68 - 72 °C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which can be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom *et al.*, *PCR Methods Applic. 1*, 111-119, 1991). In this method, multiple restriction enzyme digestions and ligations are used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

Another method which can be used to retrieve unknown sequences is that of Parker *et al.*, *Nucleic Acids Res. 19*, 3055-3060, 1991. Additionally, PCR, nested primers, and PROMOTERFINDER libraries (CLONTECH, Palo Alto, Calif.) can be used to walk genomic DNA. This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable,

in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries can be useful for extension of sequence into 5' non-transcribed regulatory regions.

Commercially available capillary electrophoresis systems can be used to analyze the size or confirm the nucleotide sequence of PCR or sequencing products. For example, capillary sequencing can employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity can be converted to electrical signal using appropriate software (*e.g.* GENOTYPER and Sequence NAVIGATOR, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display can be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

Obtaining Polypeptides

Matriptase-like serine protease polypeptides can be obtained, for example, by purification from human cells, by expression of matriptase-like serine protease polynucleotides, or by direct chemical synthesis.

Protein Purification

Matriptase-like serine protease polypeptides can be purified from human cells, such as primary tumor cells, metastatic cells, or cancer cell lines (*e.g.*, colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, or the H392 glioblastoma cell line). Carcinoma of the lung is an especially useful source of matriptase-like serine protease polypeptides. A purified matriptase-like

serine protease polypeptide is separated from other compounds which normally associate with the matriptase-like serine protease polypeptide in the cell, such as certain proteins, carbohydrates, or lipids, using methods well-known in the art. Such methods include, but are not limited to, size exclusion chromatography, ammonium sulfate fractionation, ion exchange chromatography, affinity chromatography, and preparative gel electrophoresis. A preparation of purified matriptase-like serine protease polypeptides is at least 80% pure; preferably, the preparations are 90%, 95%, or 99% pure. Purity of the preparations can be assessed by any means known in the art, such as SDS-polyacrylamide gel electrophoresis. Enzymatic activity of the purified preparations can be assayed, for example, as described in Lin *et al.*, *J. Biol. Chem. 274*, 18231-36, 1999.

Expression of Polynucleotides

To express a matriptase-like serine protease polypeptide, a matriptase-like serine protease polynucleotide can be inserted into an expression vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art can be used to construct expression vectors containing sequences encoding matriptase-like serine protease polypeptides and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook *et al.* (1989) and Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y, 1989.

A variety of expression vector/host systems can be utilized to contain and express sequences encoding a matriptase-like serine protease polypeptide. These include, but are not limited to, microorganisms, such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors, insect cell systems infected with virus expression vectors (*e.g.*, baculovirus), plant cell systems transformed with virus expression vectors (*e.g.*,

cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids), or animal cell systems.

The control elements or regulatory sequences are those non-translated regions of the vector -- enhancers, promoters, 5' and 3' untranslated regions -- which interact with host cellular proteins to carry out transcription and translation. Such elements can vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, can be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, LaJolla, Calif.) or pSPORT1 plasmid (Life Technologies) and the like can be used. The baculovirus polyhedrin promoter can be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (*e.g.*, heat shock, RUBISCO, and storage protein genes) or from plant viruses (*e.g.*, viral promoters or leader sequences) can be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of a nucleotide sequence encoding a matriptase-like serine protease polypeptide, vectors based on SV40 or EBV can be used with an appropriate selectable marker.

Bacterial and Yeast Expression Systems

In bacterial systems, a number of expression vectors can be selected depending upon the use intended for the matriptase-like serine protease polypeptide. For example, when a large quantity of a matriptase-like serine protease polypeptide is needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified can be used. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the matriptase-like serine protease polypeptide can be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid

protein is produced. pIN vectors (Van Heeke & Schuster, *J. Biol. Chem. 264*, 5503-5509, 1989 or pGEX vectors (Promega, Madison, Wis.) can be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems can be designed to include heparin, thrombin, or Factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH can be used. For reviews, see Ausubel *et al.* (1989) and Grant *et al.*, *Methods Enzymol. 153*, 516-544, 1987.

Plant and Insect Expression Systems

If plant expression vectors are used, the expression of sequences encoding matriptase-like serine protease polypeptides can be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV can be used alone or in combination with the omega leader sequence from TMV (Takamatsu *EMBO J. 6*, 307-311, 1987). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters can be used (Coruzzi *et al.*, *EMBO J. 3*, 1671-1680, 1984; Broglie *et al.*, *Science 224*, 838-843, 1984; Winter *et al.*, *Results Probl. Cell Differ. 17*, 85-105, 1991). These constructs can be introduced into plant cells by direct DNA transformation or by pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs or Murray, in MCGRAW HILL YEARBOOK OF SCIENCE AND TECHNOLOGY, McGraw Hill, New York, N.Y., pp. 191-196, 1992).

An insect system also can be used to express a matriptase-like serine protease polypeptide. For example, in one such system *Autographa californica* nuclear poly-

hedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. Sequences encoding matriptase-like serine protease polypeptides can be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of matriptase-like serine protease polypeptides will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses can then be used to infect, for example, *S. frugiperda cells* or *Trichoplusia* larvae in which matriptase-like serine protease polypeptides can be expressed (Engelhard *et al.*, *Proc. Nat. Acad. Sci. 91*, 3224-3227, 1994).

Mammalian Expression Systems

A number of viral-based expression systems can be utilized in mammalian host cells. For example, if an adenovirus is used as an expression vector, sequences encoding matriptase-like serine protease polypeptides can be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome can be used to obtain a viable virus which is capable of expressing a matriptase-like serine protease polypeptide in infected host cells (Logan & Shenk, *Proc. Natl. Acad. Sci. 81*, 3655-3659, 1984). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, can be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) also can be used to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6M to 10M are constructed and delivered to cells via conventional delivery methods (*e.g.*, liposomes, polycationic amino polymers, or vesicles).

Specific initiation signals also can be used to achieve more efficient translation of sequences encoding matriptase-like serine protease polypeptides. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences

encoding a matriptase-like serine protease polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals (including the ATG initiation codon) should be provided. The initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used (see Scharf *et al.*, *Results Probl. Cell Differ. 20*, 125-162, 1994).

Host Cells

A host cell strain can be chosen for its ability to modulate the expression of the inserted sequences or to process an expressed matriptase-like serine protease polypeptide in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the polypeptide also can be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (*e.g.*, CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; 10801 University Boulevard, Manassas, VA 20110-2209) and can be chosen to ensure the correct modification and processing of the foreign protein.

Stable expression is preferred for long-term, high-yield production of recombinant proteins. For example, cell lines which stably express matriptase-like serine protease polypeptides can be transformed using expression vectors which can contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells can be allowed to grow for 1-2 days in an enriched medium before they are

switched to a selective medium. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced matriptase-like serine protease sequences. Resistant clones of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems can be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, *Cell 11*, 223-32, 1977) and adenine phosphoribosyltransferase (Lowy *et al.*, *Cell 22*, 817-23, 1980). Genes which can be employed in *tk*$^-$ or *aprt*$^-$ cells, respectively. Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate (Wigler *et al.*, *Proc. Natl. Acad. Sci. 77*, 3567-70, 1980); *npt* confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al.*, *J. Mol. Biol. 150*, 1-14, 1981); and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murray, 1992 *supra*). Additional selectable genes have been described, for example *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. 85*, 8047-51, 1988). Visible markers such as anthocyanins, β-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, can be used to identify transformants and to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes *et al.*, *Methods Mol. Biol. 55*, 121-131, 1995).

Detecting Expression of Polypeptides

Although the presence of marker gene expression suggests that the matriptase-like serine protease polynucleotide is also present, its presence and expression may need to be confirmed. For example, if a sequence encoding a matriptase-like serine protease polypeptide is inserted within a marker gene sequence, transformed cells containing sequences which encode a matriptase-like serine protease polypeptide can be

identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a matriptase-like serine protease polypeptide under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the matriptase-like serine protease polynucleotide.

Alternatively, host cells which contain a matriptase-like serine protease polynucleotide and which express a matriptase-like serine protease polypeptide can be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip-based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of a polynucleotide sequence encoding a matriptase-like serine protease polypeptide can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding a matriptase-like serine protease polypeptide. Nucleic acid amplification-based assays involve the use of oligonucleotides selected from sequences encoding a matriptase-like serine protease polypeptide to detect transformants which contain a matriptase-like serine protease polynucleotide.

A variety of protocols for detecting and measuring the expression of a matriptase-like serine protease polypeptide, using either polyclonal or monoclonal antibodies specific for the polypeptide, are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay using monoclonal antibodies reactive to two non-interfering epitopes on a matriptase-like serine protease polypeptide can be used, or a competitive binding assay can be employed. These and other assays are described in Hampton *et al.*, SEROLOGICAL METHODS: A LABORATORY MANUAL, APS Press, St. Paul, Minn., 1990) and Maddox *et al.*, *J. Exp. Med. 158*, 1211-1216, 1983).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding matriptase-like serine protease polypeptides include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, sequences encoding a matriptase-like serine protease polypeptide can be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and can be used to synthesize RNA probes *in vitro* by addition of labeled nucleotides and an appropriate RNA polymerase, such as T7, T3, or SP6. These procedures can be conducted using a variety of commercially available kits (Amersham Pharmacia Biotech, Promega, and US Biochemical). Suitable reporter molecules or labels which can be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Expression and Purification of Polypeptides

Host cells transformed with nucleotide sequences encoding a matriptase-like serine protease polypeptide can be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The polypeptide produced by a transformed cell can be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode matriptase-like serine protease polypeptides can be designed to contain signal sequences which direct secretion of matriptase-like serine protease polypeptides through a prokaryotic or eukaryotic cell membrane.

Other constructions can be used to join a sequence encoding a matriptase-like serine protease polypeptide to a nucleotide sequence encoding a polypeptide domain which

will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor Xa or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the matriptase-like serine protease polypeptide can be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a matriptase-like serine protease polypeptide and 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography as described in Porath *et al.*, *Prot. Exp. Purif. 3*, 263-281, 1992), while the enterokinase cleavage site provides a means for purifying the matriptase-like serine protease polypeptide from the fusion protein. Vectors which contain fusion proteins are disclosed in Kroll *et al.*, *DNA Cell Biol. 12*, 441-453, 1993).

Chemical Synthesis

Sequences encoding a matriptase-like serine protease polypeptide can be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers *et al.*, *Nucl. Acids Res. Symp. Ser.* 215-223, 1980; Horn *et al. Nucl. Acids Res. Symp. Ser.* 225-232, 1980). Alternatively, a matriptase-like serine protease polypeptide itself can be produced using chemical methods to synthesize its amino acid sequence. For example, matriptase-like serine protease polypeptides can be produced by direct peptide synthesis using solid-phase techniques (Merrifield, *J. Am. Chem. Soc. 85*, 2149-2154, 1963; Roberge *et al.*, *Science 269*, 202-204, 1995). Protein synthesis can be performed using manual techniques or by automation. Automated synthesis can be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of matriptase-like serine protease polypeptides

can be separately synthesized and combined using chemical methods to produce a full-length molecule.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, PROTEINS: STRUCTURES AND MOLECULAR PRINCIPLES, WH Freeman and Co., New York, N.Y., 1983). The composition of a synthetic matriptase-like serine protease polypeptide can be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure; *see* Creighton, *supra*). Additionally, any portion of the amino acid sequence of the matriptase-like serine protease polypeptide can be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins to produce a variant polypeptide or a fusion protein.

Production of Altered Polypeptides

As will be understood by those of skill in the art, it may be advantageous to produce matriptase-like serine protease polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences disclosed herein can be engineered using methods generally known in the art to alter matriptase-like serine protease polypeptide-encoding sequences for a variety of reasons, including modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides can be used to engineer the nucleotide sequences. For example, site-directed mutagenesis can be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

Antibodies

Any type of antibody known in the art can be generated to bind specifically to an epitope of a matriptase-like serine protease polypeptide. "Antibody" as used herein includes intact immunoglobulin molecules, as well as fragments thereof, such as Fab, $F(ab')_2$, and Fv, which are capable of binding an epitope of a matriptase-like serine protease polypeptide. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. However, epitopes which involve non-contiguous amino acids may require more, *e.g.*, at least 15, 25, or 50 amino acids.

An antibody which specifically binds to an epitope of a matriptase-like serine protease polypeptide can be used therapeutically, as well as in immunochemical assays, including but not limited to Western blots, ELISAs, radioimmunoassays, immunohistochemical assays, immunoprecipitations, or other immunochemical assays known in the art. Various immunoassays can be used to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays are well known in the art. Such immunoassays typically involve the measurement of complex formation between an immunogen and an antibody which specifically binds to the immunogen.

Typically, an antibody which specifically binds to a matriptase-like serine protease polypeptide provides a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in an immunochemical assay. Preferably, antibodies which specifically bind to matriptase-like serine protease polypeptides do not detect other proteins in immunochemical assays and can immunoprecipitate a matriptase-like serine protease polypeptide from solution.

Matriptase-like serine protease polypeptides can be used to immunize a mammal, such as a mouse, rat, rabbit, guinea pig, monkey, or human, to produce polyclonal antibodies. If desired, a matriptase-like serine protease polypeptide can be conju-

gated to a carrier protein, such as bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin. Depending on the host species, various adjuvants can be used to increase the immunological response. Such adjuvants include, but are not limited to, Freund's adjuvant, mineral gels (*e.g.*, aluminum hydroxide), and surface active substances (*e.g.* lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol). Among adjuvants used in humans, BCG (*bacilli Calmette-Guerin*) and *Corynebacterium parvum* are especially useful.

Monoclonal antibodies which specifically bind to a matriptase-like serine protease polypeptide can be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These techniques include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler *et al.*, *Nature 256*, 495-497, 1985; Kozbor *et al.*, *J. Immunol. Methods 81*, 31-42, 1985; Cote *et al.*, *Proc. Natl. Acad. Sci. 80*, 2026-2030, 1983; Cole *et al.*, *Mol. Cell Biol. 62*, 109-120, 1984).

In addition, techniques developed for the production of "chimeric antibodies," the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison *et al.*, *Proc. Natl. Acad. Sci. 81*, 6851-6855, 1984; Neuberger *et al.*, *Nature 312*, 604-608, 1984; Takeda *et al.*, *Nature 314*, 452-454, 1985). Monoclonal and other antibodies also can be "humanized" to prevent a patient from mounting an immune response against the antibody when it is used therapeutically. Such antibodies may be sufficiently similar in sequence to human antibodies to be used directly in therapy or may require alteration of a few key residues. Sequence differences between rodent antibodies and human sequences can be minimized by replacing residues which differ from those in the human sequences by site directed mutagenesis of individual residues or by grating of entire complementarity determining regions. Alternatively, one can produce humanized antibodies using recombinant methods, as described in GB2188638B. Antibodies which specifically bind to a matriptase-like serine prote-

ase polypeptide can contain antigen binding sites which are either partially or fully humanized, as disclosed in U.S. 5,565,332.

Alternatively, techniques described for the production of single chain antibodies can be adapted using methods known in the art to produce single chain antibodies which specifically bind to matriptase-like serine protease polypeptides. Antibodies with related specificity, but of distinct idiotypic composition, can be generated by chain shuffling from random combinatorial immunoglobin libraries (Burton, *Proc. Natl. Acad. Sci. 88*, 11120-23, 1991).

Single-chain antibodies also can be constructed using a DNA amplification method, such as PCR, using hybridoma cDNA as a template (Thirion *et al.*, 1996, *Eur. J. Cancer Prev. 5*, 507-11). Single-chain antibodies can be mono- or bispecific, and can be bivalent or tetravalent. Construction of tetravalent, bispecific single-chain antibodies is taught, for example, in Coloma & Morrison, 1997, *Nat. Biotechnol. 15*, 159-63. Construction of bivalent, bispecific single-chain antibodies is taught in Mallender & Voss, 1994, *J. Biol. Chem. 269*, 199-206.

A nucleotide sequence encoding a single-chain antibody can be constructed using manual or automated nucleotide synthesis, cloned into an expression construct using standard recombinant DNA methods, and introduced into a cell to express the coding sequence, as described below. Alternatively, single-chain antibodies can be produced directly using, for example, filamentous phage technology. Verhaar *et al.*, 1995, *Int. J. Cancer 61*, 497-501; Nicholls *et al.*, 1993, *J. Immunol. Meth. 165*, 81-91.

Antibodies which specifically bind to matriptase-like serine protease polypeptides also can be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi *et al.*, *Proc. Natl. Acad. Sci. 86*, 3833-3837, 1989; Winter *et al.*, *Nature 349*, 293-299, 1991).

Other types of antibodies can be constructed and used therapeutically in methods of the invention. For example, chimeric antibodies can be constructed as disclosed in WO 93/03151. Binding proteins which are derived from immunoglobulins and which are multivalent and multispecific, such as the "diabodies" described in WO 94/13804, also can be prepared.

Antibodies of the invention can be purified by methods well known in the art. For example, antibodies can be affinity purified by passage over a column to which a matriptase-like serine protease polypeptide is bound. The bound antibodies can then be eluted from the column using a buffer with a high salt concentration.

Antisense Oligonucleotides

Antisense oligonucleotides are nucleotide sequences which are complementary to a specific DNA or RNA sequence. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form complexes and block either transcription or translation. Preferably, an antisense oligonucleotide is at least 11 nucleotides in length, but can be at least 12, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides long. Longer sequences also can be used. Antisense oligonucleotide molecules can be provided in a DNA construct and introduced into a cell as described above to decrease the level of matriptase-like serine protease gene products in the cell.

Antisense oligonucleotides can be deoxyribonucleotides, ribonucleotides, or a combination of both. Oligonucleotides can be synthesized manually or by an automated synthesizer, by covalently linking the 5' end of one nucleotide with the 3' end of another nucleotide with non-phosphodiester internucleotide linkages such alkylphosphonates, phosphorothioates, phosphorodithioates, alkylphosphonothioates, alkylphosphonates, phosphoramidates, phosphate esters, carbamates, acetamidate, carboxymethyl esters, carbonates, and phosphate triesters. *See* Brown, *Meth. Mol. Biol.*

20, 1-8, 1994; Sonveaux, *Meth. Mol. Biol. 26*, 1-72, 1994; Uhlmann *et al.*, *Chem. Rev. 90*, 543-583, 1990.

Modifications of matriptase-like serine protease gene expression can be obtained by designing antisense oligonucleotides which will form duplexes to the control, 5', or regulatory regions of the matriptase-like serine protease gene. Oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or chaperons. Therapeutic advances using triplex DNA have been described in the literature (*e.g.*, Gee *et al.*, in Huber & Carr, MOLECULAR AND IMMUNOLOGIC APPROACHES, Futura Publishing Co., Mt. Kisco, N.Y., 1994). An antisense oligonucleotide also can be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Precise complementarity is not required for successful duplex formation between an antisense oligonucleotide and the complementary sequence of a matriptase-like serine protease polynucleotide. Antisense oligonucleotides which comprise, for example, 2, 3, 4, or 5 or more stretches of contiguous nucleotides which are precisely complementary to a matriptase-like serine protease polynucleotide, each separated by a stretch of contiguous nucleotides which are not complementary to adjacent matriptase-like serine protease nucleotides, can provide targeting specificity for matriptase-like serine protease mRNA. Preferably, each stretch of complementary contiguous nucleotides is at least 4, 5, 6, 7, or 8 or more nucleotides in length. Non-complementary intervening sequences are preferably 1, 2, 3, or 4 nucleotides in length. One skilled in the art can easily use the calculated melting point of an antisense-sense pair to determine the degree of mismatching which will be tolerated between a particular antisense oligonucleotide and a particular matriptase-like serine protease polynucleotide sequence.

Antisense oligonucleotides can be modified without affecting their ability to hybridize to a matriptase-like serine protease polynucleotide. These modifications can be internal or at one or both ends of the antisense molecule. For example, internucleoside phosphate linkages can be modified by adding cholesteryl or diamine moieties with varying numbers of carbon residues between the amino groups and terminal ribose. Modified bases and/or sugars, such as arabinose instead of ribose, or a 3', 5'-substituted oligonucleotide in which the 3' hydroxyl group or the 5' phosphate group are substituted, also can be employed in a modified antisense oligonucleotide. These modified oligonucleotides can be prepared by methods well known in the art. *See, e.g.*, Agrawal *et al.*, *Trends Biotechnol. 10*, 152-158, 1992; Uhlmann *et al.*, *Chem. Rev. 90*, 543-584, 1990; Uhlmann *et al.*, *Tetrahedron. Lett. 215*, 3539-3542, 1987.

Ribozymes

Ribozymes are RNA molecules with catalytic activity. *See, e.g.*, Cech, *Science 236*, 1532-1539; 1987; Cech, *Ann. Rev. Biochem. 59*, 543-568; 1990, Cech, *Curr. Opin. Struct. Biol. 2*, 605-609; 1992, Couture & Stinchcomb, *Trends Genet. 12*, 510-515, 1996. Ribozymes can be used to inhibit gene function by cleaving an RNA sequence, as is known in the art (*e.g.*, Haseloff *et al.*, U.S. Patent 5,641,673). The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of specific nucleotide sequences.

The coding sequence of a matriptase-like serine protease polynucleotide can be used to generate ribozymes which will specifically bind to mRNA transcribed from the matriptase-like serine protease polynucleotide. Methods of designing and constructing ribozymes which can cleave other RNA molecules in trans in a highly sequence specific manner have been developed and described in the art (*see* Haseloff *et al.* *Nature 334*, 585-591, 1988). For example, the cleavage activity of ribozymes can be

targeted to specific RNAs by engineering a discrete "hybridization" region into the ribozyme. The hybridization region contains a sequence complementary to the target RNA and thus specifically hybridizes with the target (see, for example, Gerlach *et al.*, EP 321,201).

Specific ribozyme cleavage sites within a matriptase-like serine protease RNA target are initially identified by scanning the RNA molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the matriptase-like serine protease target RNA containing the cleavage site can be evaluated for secondary structural features which may render the target inoperable. The suitability of candidate targets also can be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays. Longer complementary sequences can be used to increase the affinity of the hybridization sequence for the target. The hybridizing and cleavage regions of the ribozyme can be integrally related; thus, upon hybridizing to the matriptase-like serine protease target RNA through the complementary regions, the catalytic region of the ribozyme can cleave the target.

Ribozymes can be introduced into cells as part of a DNA construct. Mechanical methods, such as microinjection, liposome-mediated transfection, electroporation, or calcium phosphate precipitation, can be used to introduce a ribozyme-containing DNA construct into cells in which it is desired to decrease matriptase-like serine protease expression. Alternatively, if it is desired that the cells stably retain the DNA construct, it can be supplied on a plasmid and maintained as a separate element or integrated into the genome of the cells, as is known in the art. The DNA construct can include transcriptional regulatory elements, such as a promoter element, an enhancer or UAS element, and a transcriptional terminator signal, for controlling transcription of ribozymes in the cells.

As taught in Haseloff *et al.*, U.S. Patent 5,641,673, ribozymes can be engineered so that ribozyme expression will occur in response to factors which induce expression of a target gene. Ribozymes also can be engineered to provide an additional level of regulation, so that destruction of matriptase-like serine protease mRNA occurs only when both a ribozyme and a target gene are induced in the cells.

Differentially Expressed Genes

Described herein are methods for the identification of genes whose products interact with human matriptase-like serine protease. Such genes may represent genes that are differentially expressed in disorders including, but not limited to, cancer, COPD, disorders of the peripheral or central nervous system, and cardiovascular diseases. Further, such genes may represent genes that are differentially regulated in response to manipulations relevant to the progression or treatment of such diseases. Additionally, such genes may have a temporally modulated expression, increased or decreased at different stages of tissue or organism development. A differentially expressed gene may also have its expression modulated under control versus experimental conditions. In addition, the human matriptase-like serine protease gene or gene product may itself be tested for differential expression.

The degree to which expression differs in a normal versus a diseased state need only be large enough to be visualized via standard characterization techniques such as differential display techniques. Other such standard characterization techniques by which expression differences may be visualized include but are not limited to, quantitative RT (reverse transcriptase), PCR, and Northern analysis.

Identification of Differentially Expressed Genes

To identify differentially expressed genes total RNA or, preferably, mRNA is isolated from tissues of interest. For example, RNA samples are obtained from tissues of experimental subjects and from corresponding tissues of control subjects. Any

RNA isolation technique that does not select against the isolation of mRNA may be utilized for the purification of such RNA samples. See, for example, Ausubel *et al.*, ed., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, Inc. New York, 1987-1993. Large numbers of tissue samples may readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski, U.S. Patent 4,843,155.

Transcripts within the collected RNA samples that represent RNA produced by differentially expressed genes are identified by methods well known to those of skill in the art. They include, for example, differential screening (Tedder *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 85*, 208-12, 1988), subtractive hybridization (Hedrick *et al.*, *Nature 308*, 149-53; Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 88*, 2825, 1984), and differential display (Liang & Pardee, *Science 257*, 967-71, 1992; U.S. Patent 5,262,311), and microarrays.

The differential expression information may itself suggest relevant methods for the treatment of disorders involving the human matriptase-like serine protease. For example, treatment may include a modulation of expression of the differentially expressed genes and/or the gene encoding the human matriptase-like serine protease. The differential expression information may indicate whether the expression or activity of the differentially expressed gene or gene product or the human matriptase-like serine protease gene or gene product are up-regulated or down-regulated.

Screening Methods

The invention provides methods for identifying modulators, *i.e.*, candidate or test compounds which bind to matriptase-like serine protease polypeptides or polynucleotides and/or have a stimulatory or inhibitory effect on, for example, expression or activity of the matriptase-like serine protease polypeptide or polynucleotide, so as to regulate degradation of the extracellular matrix. Decreased extracellular matrix degradation is useful for preventing or suppressing malignant cells from metastasizing.

Increased extracellular matrix degradation may be desired, for example, in developmental disorders characterized by inappropriately low levels of extracellular matrix degradation or in regeneration.

The invention provides assays for screening test compounds which bind to or modulate the activity of a matriptase-like serine protease polypeptide or a matriptase-like serine protease polynucleotide. A test compound preferably binds to a matriptase-like serine protease polypeptide or polynucleotide. More preferably, a test compound decreases a matriptase-like serine protease activity of a matriptase-like serine protease polypeptide or expression of a matriptase-like serine protease polynucleotide by at least about 10, preferably about 50, more preferably about 75, 90, or 100% relative to the absence of the test compound.

Test Compounds

Test compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity. Such compounds also may include, but are not limited to, other cellular proteins, peptides such as, for example, soluble peptides, including but not limited to, Ig-tailed fusion peptides, comprising extracellular portions of target gene product transmembrane receptors, and members of random peptide libraries (Lam, *et al.*, *Nature 354*, 82-84, 1991; Houghten *et al.*, *Nature 354*, 84-86, 1991), made of D- and/or L-configuration amino acids, phosphopeptides (including, but not limited to members of random or partially degenerate, directed phosphopeptide libraries (Songyang *et al.*, *Cell 72*, 767-78, 1993), antibodies (including, but not limited to, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, $F(ab')_2$ and FAb expression library fragments, and epitope-binding fragments thereof), and small organic or inorganic molecules.

The compounds can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, and can be produced recombi-

nantly, or synthesized by chemical methods known in the art. If desired, test compounds can be obtained using any of the numerous combinatorial library methods known in the art, including but not limited to, biological libraries, spatially addressable parallel solid phase or solution phase libraries, synthetic library methods requiring deconvolution, the "one-bead one-compound" library method, and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer, or small molecule libraries of compounds. *See* Lam, *Anticancer Drug Des. 12*, 145, 1997.

Methods for the synthesis of molecular libraries are well known in the art (*see*, for example, DeWitt *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 90*, 6909, 1993; Erb *et al. Proc. Natl. Acad. Sci. U.S.A. 91*, 11422, 1994; Zuckermann *et al.*, *J. Med. Chem. 37*, 2678, 1994; Cho *et al.*, *Science 261*, 1303, 1993; Carell *et al.*, *Angew. Chem. Int. Ed. Engl. 33*, 2059, 1994; Carell *et al.*, *Angew. Chem. Int. Ed. Engl. 33*, 2061; Gallop *et al.*, *J. Med. Chem. 37*, 1233, 1994). Libraries of compounds can be presented in solution (*see*, *e.g.*, Houghten, *Biotechniques 13*, 412-421, 1992), or on beads (Lam, *Nature 354*, 82-84, 1991), chips (Fodor, *Nature 364*, 555-556, 1993), bacteria or spores (Ladner, U.S. Patent 5,223,409), plasmids (Cull *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 89*, 1865-1869, 1992), or phage (Scott & Smith, *Science 249*, 386-390, 1990; Devlin, *Science 249*, 404-406, 1990); Cwirla *et al.*, *Proc. Natl. Acad. Sci. 97*, 6378-6382, 1990; Felici, *J. Mol. Biol. 222*, 301-310, 1991; and Ladner, U.S. Patent 5,223,409).

High Throughput Screening

Test compounds can be screened for the ability to bind to matriptase-like serine protease polypeptides or polynucleotides or to affect matriptase-like serine protease activity or matriptase-like serine protease gene expression using high throughput screening. Using high throughput screening, many discrete compounds can be tested in parallel so that large numbers of test compounds can be quickly screened. The most widely established techniques utilize 96-well microtiter plates. The wells of the

microtiter plates typically require assay volumes that range from 50 to 500 µl. In addition to the plates, many instruments, materials, pipettors, robotics, plate washers, and plate readers are commercially available to fit the 96-well format.

Alternatively, "free format assays," or assays that have no physical barrier between samples, can be used. For example, an assay using pigment cells (melanocytes) in a simple homogeneous assay for combinatorial peptide libraries is described by Jayawickreme *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 19,* 1614-18 (1994). The cells are placed under agarose in petri dishes, then beads that carry combinatorial compounds are placed on the surface of the agarose. The combinatorial compounds are partially released the compounds from the beads. Active compounds can be visualized as dark pigment areas because, as the compounds diffuse locally into the gel matrix, the active compounds cause the cells to change colors.

Another example of a free format assay is described by Chelsky, "Strategies for Screening Combinatorial Libraries: Novel and Traditional Approaches," reported at the First Annual Conference of The Society for Biomolecular Screening in Philadelphia, Pa. (Nov. 7-10, 1995). Chelsky placed a simple homogenous enzyme assay for carbonic anhydrase inside an agarose gel such that the enzyme in the gel would cause a color change throughout the gel. Thereafter, beads carrying combinatorial compounds via a photolinker were placed inside the gel and the compounds were partially released by UV-light. Compounds that inhibited the enzyme were observed as local zones of inhibition having less color change.

Yet another example is described by Salmon *et al.*, *Molecular Diversity 2*, 57-63 (1996). In this example, combinatorial libraries were screened for compounds that had cytotoxic effects on cancer cells growing in agar.

Another high throughput screening method is described in Beutel *et al.*, U.S. Patent 5,976,813. In this method, test samples are placed in a porous matrix. One or more assay components are then placed within, on top of, or at the bottom of a matrix such

as a gel, a plastic sheet, a filter, or other form of easily manipulated solid support. When samples are introduced to the porous matrix they diffuse sufficiently slowly, such that the assays can be performed without the test samples running together.

Binding Assays

For binding assays, the test compound is preferably a small molecule which binds to and occupies the active site or a fibronectin domain of the matriptase-like serine protease polypeptide, thereby making the active site or fibronectin domain inaccessible to substrate such that normal biological activity is prevented. Examples of such small molecules include, but are not limited to, small peptides or peptide-like molecules. In binding assays, either the test compound or the matriptase-like serine protease polypeptide can comprise a detectable label, such as a fluorescent, radioisotopic, chemiluminescent, or enzymatic label, such as horseradish peroxidase, alkaline phosphatase, or luciferase. Detection of a test compound which is bound to the matriptase-like serine protease polypeptide can then be accomplished, for example, by direct counting of radioemmission, by scintillation counting, or by determining conversion of an appropriate substrate to a detectable product.

Alternatively, binding of a test compound to a matriptase-like serine protease polypeptide can be determined without labeling either of the interactants. For example, a microphysiometer can be used to detect binding of a test compound with a target polypeptide. A microphysiometer (*e.g.*, Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a test compound and a matriptase-like serine protease polypeptide. (McConnell *et al.*, *Science 257*, 1906-1912, 1992).

Determining the ability of a test compound to bind to a matriptase-like serine protease polypeptide also can be accomplished using a technology such as real-time Bi-

molecular Interaction Analysis (BIA). Sjolander & Urbaniczky, *Anal. Chem. 63*, 2338-2345, 1991, and Szabo *et al.*, *Curr. Opin. Struct. Biol. 5*, 699-705, 1995. BIA is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In yet another aspect of the invention, a matriptase-like serine protease polypeptide can be used as a "bait protein" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent 5,283,317; Zervos *et al.*, *Cell 72*, 223-232, 1993; Madura *et al.*, *J. Biol. Chem. 268*, 12046-12054, 1993; Bartel *et al.*, *Biotechniques 14*, 920-924, 1993; Iwabuchi *et al.*, *Oncogene 8*, 1693-1696, 1993; and Brent W094/10300), to identify other proteins which bind to or interact with the matriptase-like serine protease polypeptide and modulate its activity.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. For example, in one construct a polynucleotide encoding a matriptase-like serine protease polypeptide is fused to a polynucleotide encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence that encodes an unidentified protein ("prey" or "sample") is fused to a polynucleotide that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact *in vivo* to form an protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ), which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the DNA sequence encoding the protein which interacts with the matriptase-like serine protease polypeptide.

It may be desirable to immobilize either the matriptase-like serine protease polypeptide (or polynucleotide) or the test compound to facilitate separation of bound from unbound forms of one or both of the interactants, as well as to accommodate automation of the assay. Thus, either the matriptase-like serine protease polypeptide (or polynucleotide) or the test compound can be bound to a solid support. Suitable solid supports include, but are not limited to, glass or plastic slides, tissue culture plates, microtiter wells, tubes, silicon chips, or particles such as beads (including, but not limited to, latex, polystyrene, or glass beads). Any method known in the art can be used to attach the matriptase-like serine protease polypeptide (or polynucleotide) or test compound to a solid support, including use of covalent and non-covalent linkages, passive absorption, or pairs of binding moieties attached respectively to the polypeptide or test compound and the solid support. Test compounds are preferably bound to the solid support in an array, so that the location of individual test compounds can be tracked. Binding of a test compound to a matriptase-like serine protease polypeptide (or polynucleotide) can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and microcentrifuge tubes.

In one embodiment, a matriptase-like serine protease polypeptide is a fusion protein comprising a domain that allows the matriptase-like serine protease polypeptide to be bound to a solid support. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and the non-adsorbed matriptase-like serine protease polypeptide; the mixture is then incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components. Binding of the interactants can be determined either directly or indirectly, as described above. Alternatively, the complexes can be dissociated from the solid support before binding is determined.

Other techniques for immobilizing polypeptides or polynucleotides on a solid support also can be used in the screening assays of the invention. For example, either a matriptase-like serine protease polypeptide (or polynucleotide) or a test compound can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated matriptase-like serine protease polypeptides or test compounds can be prepared from biotin-NHS(N-hydroxysuccinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.) and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies which specifically bind to a matriptase-like serine protease polypeptide polynucleotides, or a test compound, but which do not interfere with a desired binding site, such as the active site or a fibronectin domain of the matriptase-like serine protease polypeptide, can be derivatized to the wells of the plate. Unbound target or protein can be trapped in the wells by antibody conjugation.

Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies which specifically bind to the matriptase-like serine protease polypeptide (or polynucleotides) or test compound, enzyme-linked assays which rely on detecting a matriptase-like serine protease activity of the matriptase-like serine protease polypeptide, and SDS gel electrophoresis under non-reducing conditions.

Screening for test compounds which bind to a matriptase-like serine protease polypeptide or polynucleotide also can be carried out in an intact cell. Any cell which comprises a matriptase-like serine protease polynucleotide or polypeptide can be used in a cell-based assay system. A matriptase-like serine protease polynucleotide can be naturally occurring in the cell or can be introduced using techniques such as those described above. Either a primary culture or an established cell line, including neoplastic cell lines such as the colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, and the H392 glioblas-

toma cell line, can be used. An intact cell is contacted with a test compound. Binding of the test compound to a matriptase-like serine protease polypeptide or polynucleotide is determined as described above, after lysing the cell to release the matriptase-like serine protease polypeptide-test compound complex.

Enzyme Assays

Test compounds can be tested for the ability to increase or decrease a matriptase-like serine protease activity of a matriptase-like serine protease polypeptide. Matriptase-like serine protease activity can be measured, for example, as described in Lin *et al.*, *J. Biol. Chem. 274*, 18231-36, 1999. Matriptase-like serine protease activity can be measured after contacting either a purified matriptase-like serine protease polypeptide, a cell extract, or an intact cell with a test compound. A test compound which decreases matriptase-like serine protease activity by at least about 10, preferably about 50, more preferably about 75, 90, or 100% is identified as a potential therapeutic agent for decreasing extracellular matrix degradation. A test compound which increases matriptase-like serine protease activity by at least about 10, preferably about 50, more preferably about 75, 90, or 100% is identified as a potential therapeutic agent for increasing extracellular matrix degradation.

Gene Expression

In another embodiment, test compounds which increase or decrease matriptase-like serine protease gene expression are identified. A matriptase-like serine protease polynucleotide is contacted with a test compound, and the expression of an RNA or polypeptide product of the matriptase-like serine protease polynucleotide is determined. The level of expression of matriptase-like serine protease mRNA or polypeptide in the presence of the test compound is compared to the level of expression of matriptase-like serine protease mRNA or polypeptide in the absence of the test compound. The test compound can then be identified as a modulator of expression based on this comparison. For example, when expression of matriptase-like serine

protease mRNA or polypeptide is greater in the presence of the test compound than in its absence, the test compound is identified as a stimulator or enhancer of matriptase-like serine protease mRNA or polypeptide is less expression. Alternatively, when expression of the mRNA or protein is less in the presence of the test compound than in its absence, the test compound is identified as an inhibitor of matriptase-like serine protease mRNA or polypeptide expression.

The level of matriptase-like serine protease mRNA or polypeptide expression in the cells can be determined by methods well known in the art for detecting mRNA or protein. Either qualitative or quantitative methods can be used. The presence of polypeptide products of a matriptase-like serine protease polynucleotide can be determined, for example, using a variety of techniques known in the art, including immunochemical methods such as radioimmunoassay, Western blotting, and immunohistochemistry. Alternatively, polypeptide synthesis can be determined *in vivo*, in a cell culture, or in an *in vitro* translation system by detecting incorporation of labeled amino acids into a matriptase-like serine protease polypeptide.

Such screening can be carried out either in a cell-free assay system or in an intact cell. Any cell which expresses a matriptase-like serine protease polynucleotide can be used in a cell-based assay system. The matriptase-like serine protease polynucleotide can be naturally occurring in the cell or can be introduced using techniques such as those described above. Either a primary culture or an established cell line, including neoplastic cell lines such as the colon cancer cell lines HCT116, DLD1, HT29, Caco2, SW837, SW480, and RKO, breast cancer cell lines 21-PT, 21-MT, MDA-468, SK-BR3, and BT-474, the A549 lung cancer cell line, and the H392 glioblastoma cell line, can be used.

Pharmaceutical Compositions

The invention also provides pharmaceutical compositions which can be administered to a patient to achieve a therapeutic effect. Pharmaceutical compositions of the in-

vention can comprise a matriptase-like serine protease polypeptide, matriptase-like serine protease polynucleotide, antibodies which specifically bind to a matriptase-like serine protease polypeptide, or mimetics, agonists, antagonists, or inhibitors of a matriptase-like serine protease polypeptide. The compositions can be administered alone or in combination with at least one other agent, such as stabilizing compound, which can be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions can be administered to a patient alone, or in combination with other agents, drugs or hormones.

In addition to the active ingredients, these pharmaceutical compositions can contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Pharmaceutical compositions of the invention can be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, parenteral, topical, sublingual, or rectal means. Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or

solubilizing agents can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which also can contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions can contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds can be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers also can be used for delivery. Optionally, the suspension also can contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. For topical or nasal administration, penetrants appropriate to the particular barrier to

be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention can be manufactured in a manner that is known in the art, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. The pharmaceutical composition can be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation can be a lyophilized powder which can contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

Further details on techniques for formulation and administration can be found in the latest edition of REMINGTON'S PHARMACEUTICAL SCIENCES (Maack Publishing Co., Easton, Pa.). After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. Such labeling would include amount, frequency, and method of administration.

Therapeutic Indications and Methods

1. Tumor Cell Invasion and Metastasis. The human matriptase-like serine protease gene provides a therapeutic target for decreasing extracellular matrix degradation, in particular for treating or preventing metastatic cancer. Cancer is a disease fundamentally caused by oncogenic cellular transformation. There are several hallmarks of transformed cells that distinguish them from their normal counterparts and underlie the pathophysiology of cancer. These include uncontrolled cellular proliferation, unresponsiveness to normal death-inducing signals (immortalization), increased cellular motility and invasiveness, increased ability to recruit blood supply through induction of new blood vessel formation (angiogenesis), genetic instability, and dys-

regulated gene expression. Various combinations of these aberrant physiologies, along with the acquisition of drug-resistance frequently lead to an intractable disease state in which organ failure and patient death ultimately ensue.

Most standard cancer therapies target cellular proliferation and rely on the differential proliferative capacities between transformed and normal cells for their efficacy. This approach is hindered by the facts that several important normal cell types are also highly proliferative and that cancer cells frequently become resistant to these agents. Thus, the therapeutic indices for traditional anti-cancer therapies rarely exceed 2.0.

The advent of genomics-driven molecular target identification has opened up the possibility of identifying new cancer-specific targets for therapeutic intervention that will provide safer, more effective treatments for cancer patients. Thus, newly discovered tumor-associated genes and their products can be tested for their role(s) in disease and used as tools to discover and develop innovative therapies. Genes playing important roles in any of the physiological processes outlined above can be characterized as cancer targets.

Genes or gene fragments identified through genomics can readily be expressed in one or more heterologous expression systems to produce functional recombinant proteins. These proteins are characterized *in vitro* for their biochemical properties and then used as tools in high-throughput molecular screening programs to identify chemical modulators of their biochemical activities. Agonists and/or antagonists of target protein activity can be identified in this manner and subsequently tested in cellular and *in vivo* disease models for anti-cancer activity. Optimization of lead compounds with iterative testing in biological models and detailed pharmacokinetic and toxicological analyses form the basis for drug development and subsequent testing in humans.

For example, blocking a fibronectin domain of human ephrin-like serine protease can suppress or prevent migration or metastasis of tumor cells in response to fibronectin (**9, 10**). Cancers whose metastasis can be suppressed according to the invention include adenocarcinoma, melanoma, cancers of the adrenal gland, bladder, bone, breast, cervix, gall bladder, liver, lung, ovary, pancreas, prostate, testis, and uterus. Circulating tumor cells arrested in the capillary beds of different organs must invade the endothelial cell lining and degrade its underlying basement membrane (BM) in order to invade into the extravascular tissue(s) where they establish metastasis (**1, 2**). Metastatic tumor cells often attach at or near the intercellular junctions between adjacent endothelial cells. Such attachment of the metastatic cells is followed by rupture of the junctions, retraction of the endothelial cell borders and migration through the breach in the endothelium toward the exposed underlying BM (**1, 11**).

Once located between endothelial cells and the BM, the invading cells must degrade the subendothelial glycoproteins and proteoglycans of the BM in order to migrate out of the vascular compartment. Several cellular enzymes (*e.g.*, collagenase IV, plasminogen activator, cathepsin B, elastase) are thought to be involved in degradation of BM (**2, 11**). Suppression of human matriptase-like serine protease activity therefore can be used to suppress tumor cell invasion and metastasis.

2. Tumor Angiogenesis. Basic fibroblast growth factor (bFGF) has been extracted from the subendothelial extracellular matrix produced *in vitro* (**3**) and from basement membranes of the cornea (**4**), suggesting that extracellular matrix may serve as a reservoir for bFGF. Immunohistochemical staining revealed the localization of bFGF in basement membranes of diverse tissues and blood vessels (**5**). Despite the ubiquitous presence of bFGF in normal tissues, endothelial cell proliferation in these tissues is usually very low, which suggests that bFGF is somehow sequestered from its site of action. It is possible, therefore, that suppression of human matriptase-like serine protease activity can suppress release of active bFGF from extracellular matrix and basement membranes. In addition, displacement of bFGF from its storage within basement membranes and extracellular matrix may therefore provide a novel mecha-

nism for induction of neovascularization in normal and pathological situations. Restriction of endothelial cell growth factors in the extracellular matrix may prevent their systemic action on the vascular endothelium, thus maintaining a very low rate of endothelial cells turnover and vessel growth. On the other hand, release of bFGF from storage in the extracellular matrix may elicit localized endothelial cell proliferation and neovascularization in processes such as wound healing, inflammation and tumor development (6, 7).

3. Inflammation and Cellular Immunity. Matriptase-like serine protease activity may be involved in the ability of activated cells of the immune system to leave the circulation and elicit both inflammatory and autoimmune responses. Thus, inflammation and cellular immunity may be regulated by regulating activity of matriptase-like serine protease.

4. Viral infection. Removal of the cell surface components by matriptase-like serine protease may influence the ability of viruses to attach to the cell surface. Regulation of matriptase-like serine protease may therefore be used to treat viral infections.

5. Neurodegenerative diseases. It is also possible that matriptase-like serine protease activity can be used to degrade, for example, prion protein amyloid plaques of Genstmann-Straussler Syndrome, Creutzfeldt-Jakob disease, and Scrapie. Other disorders of the peripheral or central nervous system which can be treated include brain injuries, cerebrovascular diseases and their consequences, Parkinson's disease, corticobasal degeneration, motor neuron disease, dementia, including ALS, multiple sclerosis, traumatic brain injury, stroke, post-stroke, post-traumatic brain injury, and small-vessel cerebrovascular disease. Dementias, such as Alzheimer's disease, vascular dementia, dementia with Lewy bodies, frontotemporal dementia and Parkinsonism linked to chromosome 17, frontotemporal dementias, including Pick's disease, progressive nuclear palsy, corticobasal degeneration, Huntington's disease, thalamic degeneration, Creutzfeld-Jakob dementia, HIV dementia, schizophrenia with dementia, and Korsakoff's psychosis also can be treated. Similarly, it is possi-

ble to treat cognitive-related disorders, such as mild cognitive impairment, age-associated memory impairment, age-related cognitive decline, vascular cognitive impairment, attention deficit disorders, attention deficit hyperactivity disorders, and memory disturbances in children with learning disabilities, by regulating the activity of human matriptase-like protein.

Pain that is associated with disorders of the peripheral or central nervous system also can be treated by regulating the activity of human matriptase-like serine protease. Pain which can be treated includes that associated with central nervous system disorders, such as multiple sclerosis, spinal cord injury, sciatica, failed back surgery syndrome, traumatic brain injury, epilepsy, Parkinson's disease, post-stroke, and vascular lesions in the brain and spinal cord (*e.g.*, infarct, hemorrhage, vascular malformation). Non-central neuropathic pain includes that associated with post mastectomy pain, reflex sympathetic dystrophy (RSD), trigeminal neuralgiaradioculopathy, post-surgical pain, HIV/AIDS related pain, cancer pain, metabolic neuropathies (*e.g.*, diabetic neuropathy, vasculitic neuropathy secondary to connective tissue disease), paraneoplastic polyneuropathy associated, for example, with carcinoma of lung, or leukemia, or lymphoma, or carcinoma of prostate, colon or stomach, trigeminal neuralgia, cranial neuralgias, and post-herpetic neuralgia. Pain associated with cancer and cancer treatment also can be treated, as can headache pain (for example, migraine with aura, migraine without aura, and other migraine disorders), episodic and chronic tension-type headache, tension-type like headache, cluster headache, and chronic paroxysmal hemicrania.

6. Restenosis and Atherosclerosis. Proliferation of arterial smooth muscle cells (SMCs) in response to endothelial injury and accumulation of cholesterol rich lipoproteins are basic events in the pathogenesis of atherosclerosis and restenosis (8). It is possible that matriptase-like serine protease may be involved in the catabolic pathway that may allow substantial cellular and interstitial accumulation of cholesterol rich lipoproteins. The latter pathway is expected to be highly atherogenic by promoting accumulation of apoB and apoE rich lipoproteins (*i.e.* LDL, VLDL, chy-

lomicrons), independent of feedback inhibition by the cellular sterol content. Altered levels of human matriptase-like serine protease activity therefore may inhibit both SMC proliferation and lipid accumulation and thus may halt the progression of restenosis and atherosclerosis.

Other cardiovascular diseases which can be treated include the following disorders of the heart and the vascular system: congestive heart failure, myocardial infarction, ischemic diseases of the heart, all kinds of atrial and ventricular arrhythmias, hypertensive vascular diseases and peripheral vascular diseases.

Heart failure is defined as a pathophysiologic state in which an abnormality of cardiac function is responsible for the failure of the heart to pump blood at a rate commensurate with the requirement of the metabolizing tissue. It includes all forms of pumping failure such as high-output and low-output, acute and chronic, right-sided or left-sided, systolic or diastolic, independent of the underlying cause.

Myocardial infarction (MI) is generally caused by an abrupt decrease in coronary blood flow that follows a thrombotic occlusion of a coronary artery previously narrowed by arteriosclerosis. MI prophylaxis (primary and secondary prevention) is included as well as the acute treatment of MI and the prevention of complications.

Ischemic diseases are conditions in which the coronary flow is restricted resulting in an perfusion which is inadequate to meet the myocardial requirement for oxygen. This group of diseases include stable angina, unstable angina and asymptomatic ischemia.

Arrhythmias include all forms of atrial and ventricular tachyarrhythmias (atrial tachycardia, atrial flutter, atrial fibrillation, atrio-ventricular reentrant tachycardia, preexcitation syndrome, ventricular tachycardia, ventricular flutter, ventricular fibrillation) as well as bradycardic forms of arrhythmias.

Hypertensive vascular diseases include primary as well as all kinds of secondary arterial hypertension (renal, endocrine, neurogenic, others). The genes may be used as drug targets for the treatment of hypertension as well as for the prevention of all complications. Peripheral vascular diseases are defined as vascular diseases in which arterial and/or venous flow is reduced resulting in an imbalance between blood supply and tissue oxygen demand. It includes chronic peripheral arterial occlusive disease (PAOD), acute arterial thrombosis and embolism, inflammatory vascular disorders, Raynaud's phenomenon and venous disorders.

7. COPD. Chronic obstructive pulmonary (or airways) disease (COPD) is a condition defined physiologically as airflow obstruction that generally results from a mixture of emphysema and peripheral airway obstruction due to chronic bronchitis (Senior & Shapiro, *Pulmonary Diseases and Disorders*, 3d ed., New York, McGraw-Hill, 1998, pp. 659-681, 1998; Barnes, *Chest 117*, 10S-14S, 2000). Emphysema is characterized by destruction of alveolar walls leading to abnormal enlargement of the air spaces of the lung. Chronic bronchitis is defined clinically as the presence of chronic productive cough for three months in each of two successive years. In COPD, airflow obstruction is usually progressive and is only partially reversible. By far the most important risk factor for development of COPD is cigarette smoking, although the disease does occur in non-smokers.

Chronic inflammation of the airways is a key pathological feature of COPD (Senior & Shapiro, 1998). The inflammatory cell population comprises increased numbers of macrophages, neutrophils, and $CD8^+$ lymphocyes. Inhaled irritants, such as cigarette smoke, activate macrophages which are resident in the respiratory tract, as well as epithelial cells leading to release of chemokines (*e.g.*, interleukin-8) and other chemotactic factors. These chemotactic factors act to increase the neutrophil/monocyte trafficking from the blood into the lung tissue and airways. Neutrophils and monocytes recruited into the airways can release a variety of potentially damaging mediators such as proteolytic enzymes and reactive oxygen species. Matrix degradation and emphysema, along with airway wall thickening, surfactant dys-

function, and mucus hypersecretion, all are potential sequelae of this inflammatory response that lead to impaired airflow and gas exchange.

COPD is characterized by damage to the lung extracellular matrix and emphysema can be viewed as the pathologic process that affects the lung parenchyma. This process eventually leads to the destruction of the airway walls resulting in permanent airspace enlargement (Senior and Shapiro, in PULMONARY DISEASES AND DISORDERS, 3rd ed., New York, McGraw-Hill, 1998, pp. 659 – 681, 1998). The observation that inherited deficiency of a1-antitrypsin (a1-AT), the primary inhibitor of neutrophil elastase, predisposes individuals to early onset emphysema, and that intrapulmonary instillation of elastolytic enzymes in experimental animals causes emphysema, led to the elastase:antielastase hypothesis for the pathogenesis of emphysema (Eriksson, *Acta Med. Scand. 177(Suppl.)*, 432, 1965, Gross, *J. Occup. Med.* 6, 481-84, 1964). This in turn led to the concept that destruction of elastin in the lung parenchyma is the basis of the development of emphysema.

A broad range of immune and inflammatory cells including neutrophils, macrophages, T lymphocytes and eosinophils contain proteolytic enzymes that could contribute to the destruction of lung extracellular matrix (Shapiro, 1999). In addition, a number of different classes of proteases have been identified that have the potential to contribute to lung matrix destruction. These include serine proteases, matrix metalloproteinases and cysteine proteases. Of these classes of enzymes, a number can hydrolyze elastin and have been shown to be elevated in COPD patients (neutrophil elastase, MMP-2, 9, 12) (Culpitt *et al.*, *Am. J. Respir. Crit. Care Med. 160*, 1635-39, 1999, Shapiro, *Am. J. Crit. Care Med. 160 (5)*, S29 – S32,1999).

It is expected that in the future novel members of the existing classes of proteases and new classes of proteases will be identified that play a significant role in the damage of the extracellular lung matrix including elastin proteolysis. Novel protease targets therefore remain very attractive therapeutic targets.

8. Other therapeutic and diagnostic indications. Anti-human matriptase-like serine protease antibodies can be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions, and renal failure in biopsy specimens, plasma samples, and body fluids. Alternatively, if desired a matriptase-like serine protease function can be supplied to a cell by introducing a matriptase-like serine protease-encoding polynucleotide into the cell.

The invention further pertains to the use of novel agents identified by the screening assays described above. Accordingly, it is within the scope of this invention to use a test compound identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a modulating agent, an antisense nucleic acid molecule, a specific antibody, ribozyme, or a polypeptide-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

A reagent which affects matriptase-like serine protease activity can be administered to a human cell, either *in vitro* or *in vivo*, to reduce matriptase-like serine protease activity. The reagent preferably binds to an expression product of a human matriptase-like serine protease gene. If the expression product is a polypeptide, the reagent is preferably an antibody. For treatment of human cells *ex vivo*, an antibody can be added to a preparation of stem cells which have been removed from the body. The cells can then be replaced in the same or another human body, with or without clonal propagation, as is known in the art.

In one embodiment, the reagent is delivered using a liposome. Preferably, the liposome is stable in the animal into which it has been administered for at least about 30 minutes, more preferably for at least about 1 hour, and even more preferably for at

least about 24 hours. A liposome comprises a lipid composition that is capable of targeting a reagent, particularly a polynucleotide, to a particular site in an animal, such as a human. Preferably, the lipid composition of the liposome is capable of targeting to a specific organ of an animal, such as the lung or liver.

A liposome useful in the present invention comprises a lipid composition that is capable of fusing with the plasma membrane of the targeted cell to deliver its contents to the cell. Preferably, the transfection efficiency of a liposome is about 0.5 μg of DNA per 16 nmole of liposome delivered to about 10^6 cells, more preferably about 1.0 μg of DNA per 16 nmol of liposome delivered to about 10^6 cells, and even more preferably about 2.0 μg of DNA per 16 nmol of liposome delivered to about 10^6 cells. Preferably, a liposome is between about 100 and 500 nm, more preferably between about 150 and 450 nm, and even more preferably between about 200 and 400 nm in diameter.

Suitable liposomes for use in the present invention include those liposomes standardly used in, for example, gene delivery methods known to those of skill in the art. More preferred liposomes include liposomes having a polycationic lipid composition and/or liposomes having a cholesterol backbone conjugated to polyethylene glycol. Optionally, a liposome comprises a compound capable of targeting the liposome to a tumor cell, such as a tumor cell ligand exposed on the outer surface of the liposome.

Complexing a liposome with a reagent such as an antisense oligonucleotide or ribozyme can be achieved using methods which are standard in the art (see, for example, U.S. Patent 5,705,151). Preferably, from about 0.1 μg to about 10 μg of polynucleotide is combined with about 8 nmol of liposomes, more preferably from about 0.5 μg to about 5 μg of polynucleotides are combined with about 8 nmol liposomes, and even more preferably about 1.0 μg of polynucleotides is combined with about 8 nmol liposomes.

In another embodiment, antibodies can be delivered to specific tissues *in vivo* using receptor-mediated targeted delivery. Receptor-mediated DNA delivery techniques are taught in, for example, Findeis *et al. Trends in Biotechnol. 11*, 202-05 (1993); Chiou *et al.*, GENE THERAPEUTICS: METHODS AND APPLICATIONS OF DIRECT GENE TRANSFER (J.A. Wolff, ed.) (1994); Wu & Wu, *J. Biol. Chem. 263*, 621-24 (1988); Wu *et al.*, *J. Biol. Chem. 269*, 542-46 (1994); Zenke *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 87*, 3655-59 (1990); Wu *et al.*, *J. Biol. Chem. 266*, 338-42 (1991).

If the reagent is a single-chain antibody, polynucleotides encoding the antibody can be constructed and introduced into a cell either *ex vivo* or *in vivo* using well-established techniques including, but not limited to, transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and DEAE- or calcium phosphate-mediated transfection.

Determination of a Therapeutically Effective Dose

The determination of a therapeutically effective dose is well within the capability of those skilled in the art. A therapeutically effective dose refers to that amount of active ingredient which increases or decreases extracellular matrix degradation relative to that which occurs in the absence of the therapeutically effective dose.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in animal models, usually mice, rabbits, dogs, or pigs. The animal model also can be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

Therapeutic efficacy and toxicity, *e.g.*, ED_{50} (the dose therapeutically effective in 50% of the population) and LD_{50} (the dose lethal to 50% of the population), can be

determined by standard pharmaceutical procedures in cell cultures or experimental animals. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the ratio, LD_{50}/ED_{50}.

Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active ingredient or to maintain the desired effect. Factors which can be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts can vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Effective *in vivo* dosages of an antibody are in the range of about 5 µg to about 50 µg/kg, about 50 µg to about 5 mg/kg, about 100 µg to about 500 µg/kg of patient

body weight, and about 200 to about 250 μg/kg of patient body weight. For administration of polynucleotides encoding single-chain antibodies, effective *in vivo* dosages are in the range of about 100 ng to about 200 ng, 500 ng to about 50 mg, about 1 μg to about 2 mg, about 5 μg to about 500 μg, and about 20 μg to about 100 μg of DNA.

If the expression product is mRNA, the reagent is preferably an antisense oligonucleotide or a ribozyme. Polynucleotides which express antisense oligonucleotides or ribozymes can be introduced into cells by a variety of methods, as described above.

Preferably, a reagent reduces expression of a matriptase-like serine protease polynucleotide or activity of a matriptase-like serine protease polypeptide by at least about 10, preferably about 50, more preferably about 75, 90, or 100% relative to the absence of the reagent. The effectiveness of the mechanism chosen to decrease the level of expression of a matriptase-like serine protease polynucleotide or the activity of a matriptase-like serine protease polypeptide can be assessed using methods well known in the art, such as hybridization of nucleotide probes to matriptase-like serine protease-specific mRNA, quantitative RT-PCR, immunologic detection of a matriptase-like serine protease polypeptide, or measurement of matriptase-like serine protease activity.

In any of the embodiments described above, any of the pharmaceutical compositions of the invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy can be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents can act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Any of the therapeutic methods described above can be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

The above disclosure generally describes the present invention, and all patents and patent applications cited in this disclosure are expressly incorporated herein. A more complete understanding can be obtained by reference to the following specific examples which are provided for purposes of illustration only and are not intended to limit the scope of the invention.

EXAMPLE 1

Detection of a matriptase-like serine protease activity

The polynucleotide of SEQ ID NO: 1 or 17 is inserted into the expression vector pCEV4 and the expression vector pCEV4- matriptase-like serine protease polypeptide obtained is transfected into human embryonic kidney 293 cells.

Protease activity of cellular extracts from the transfacted cells are measured using thiobenzylester substrates, as described in U.S. Patent 5,500,344. For monitoring enzyme activities from granules and column fractions, assays are performed at room temperature using 0.5 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) (Sigma) to detect the HSBzl leaving group (ε_{410} =13600 M^{-1} cm^{-1}).

Furthermore, BLT-esterase activity is estimated using a microtiter assay (Green and Shaw, *Anal. Biochem. 93*, 223-226, 1979). Briefly, 50 µl of sample is added to 100 µl of 1 mM DTNB, made up in 10 mM HEPES, 1 mM $CaCl_2$, 1 mM $MgCl_2$, pH 7.2. The reaction is initiated by the addition of 50 µl of BLT (Sigma) to give a final concentration of 500 µM. For Metase determinations, 50 µl of dilutions of the sample in 0.1 M HEPES, 0.05 M $CaCl_2$, pH 7.5, are added to 100 µl of 1 mM DTNB, and the reaction is initiated by the addition of 50 µl of Boc-Ala-Ala-Met-S Benzyl (Bzl) to give a final concentration of 150 µM. The duration of the assay depends on color development, the rate of which is measured ($O.D._{410}$) on a Dynatech MR 5000 microplate reader. Controls of sample and DTNB alone or DTNB and substrate alone are run.

Additionally, peptide thiobenzyl ester substrates are used to measure protease activities. The chymase substrate Suc-Phe-Leu-Phe-SBzl is purchased from BACHEM Bioscience Inc., Philadelphia, Pa. Z-Arg-SBzl (the tryptase substrate, Kam *et al.*, *J. Biol. Chem. 262*, 3444-3451, 1987); Boc-Ala-Ala-AA-SBzl (AA=Asp, Met, Leu, Nle, or Ser), and Suc-Ala-Ala-Met-SBzl (Odake *et al, Biochemistry 30*, 2217-2227, 1991); Harper *et al.*, *Biochemistry 23*, 2995-3002, 1984) are synthesized previously.

Boc-Ala-Ala-Asp-SBzl is the substrate for Asp-ase and peptide thiobenzyl esters containing Met, Leu or Nle are substrates for Met-ase SP. Assays are performed at room temperature in 0.1 M, HEPES buffer, pH 7.5, containing 0.01 M $CaCl_2$ and 8% Me_2O using 0.34 mM 4,4'-dithiodipyridine (Aldrithiol-4, Aldrich Chemical Co., Milwaukee, Wis.) to detect HSBzl leaving group that reacts with 4,4'-dithiodipyridine to release thiopyridone (324=19800 M^{-1} cm^{-1}, Grasetti and Murray, *Arch. Biochem. Biophys. 119*, 41-49, 1967). The initial rates are measured at 324 nm using a Beckman 35 spectrophotometer when 10-25 µl of an enzyme stock solution is added to a cuvette containing 2.0 ml of buffer, 150 µl of 4,4'-dithiodipyridine, and 25 µl of substrate. The same volume of substrate and 4,4'-dithiodipyridine are added to the reference cell in order to compensate for the background hydrolysis rate of the substrates. Initial rates are measured in duplicate for each substrate concentration and are averaged in each case. Substrate concentrations are 100-133 µM. The matriptase-like serine protease activity of the polypeptide of SEQ ID NO: 2 or 18 is shown.

EXAMPLE 2

Identification of a test compound which binds to a matriptase-like serine protease polypeptide

Purified matriptase-like serine protease polypeptides comprising a glutathione-S-transferase protein and absorbed onto glutathione-derivatized wells of 96-well microtiter plates are contacted with test compounds from a small molecule library at pH 7.0 in a physiological buffer solution. Matriptase-like serine protease polypeptides comprise the amino acid sequence shown in SEQ ID NO:2 or 18. The test compounds comprise a fluorescent tag. The samples are incubated for 5 minutes to one hour. Control samples are incubated in the absence of a test compound.

The buffer solution containing the test compounds is washed from the wells. Binding of a test compound to a matriptase-like serine protease polypeptide is detected by fluorescence measurements of the contents of the wells. A test compound which increases the fluorescence in a well by at least 15% relative to fluorescence of a well

in which a test compound was not incubated is identified as a compound which binds to a matriptase-like serine protease polypeptide.

EXAMPLE 3

Identification of a test compound which decreases matriptase-like serine protease activity

Cellular extracts from the human colon cancer cell line HCT116 are contacted with test compounds from a small molecule library and assayed for matriptase-like serine protease activity. Control extracts, in the absence of a test compound, also are assayed. Matriptase activity is measured as described in Lin *et al.*, *J. Biol. Chem. 274*, 18231-36, 1999. A test compound which decreases matriptase-like serine protease activity of the extract relative to the control extract by at least 20% is identified as a matriptase-like serine protease inhibitor.

EXAMPLE 4

Identification of a test compound which decreases matriptase-like serine protease gene expression

A test compound is administered to a culture of the breast tumor cell line MDA-468 and incubated at 37 °C for 10 to 45 minutes. A culture of the same type of cells incubated for the same time without the test compound provides a negative control.

RNA is isolated from the two cultures as described in Chirgwin *et al.*, *Biochem. 18*, 5294-99, 1979). Northern blots are prepared using 20 to 30 µg total RNA and hybridized with a ^{32}P-labeled matriptase-like serine protease-specific probe at 65 °C in Express-hyb (CLONTECH). The probe comprises at least 11 contiguous nucleotides selected from the complement of SEQ ID NO:1 or 17. A test compound which decreases the matriptase-like serine protease -specific signal relative to the signal obtained in the absence of the test compound is identified as an inhibitor of matriptase-like serine protease gene expression.

EXAMPLE 5

Treatment of a breast tumor with a reagent which specifically binds to a matriptase-like serine protease gene product

Synthesis of antisense matriptase-like serine protease oligonucleotides comprising at least 11 contiguous nucleotides selected from the complement of SEQ ID NO:1 or 17 is performed on a Pharmacia Gene Assembler series synthesizer using the phosphoramidite procedure (Uhlmann *et al.*, *Chem. Rev. 90*, 534-83, 1990). Following assembly and deprotection, oligonucleotides are ethanol-precipitated twice, dried, and suspended in phosphate-buffered saline (PBS) at the desired concentration. Purity of these oligonucleotides is tested by capillary gel electrophoreses and ion exchange HPLC. Endotoxin levels in the oligonucleotide preparation are determined using the *Limulus* Amebocyte Assay (Bang, *Biol. Bull. (Woods Hole, Mass.) 105*, 361-362, 1953).

An aqueous composition containing the antisense oligonucleotides at a concentration of 0.1-100 μM is injected directly into a breast tumor with a needle. The needle is placed in the tumors and withdrawn while expressing the aqueous composition within the tumor.

The breast tumor is monitored over a period of days or weeks. Additional injections of the antisense oligonucleotides can be given during that time. Metastasis of the breast tumor is suppressed due to decreased matriptase-like serine protease activity of the breast tumor cells.

EXAMPLE 6

Proliferation inhibition assay: Antisense oligonucleotides suppress the growth of cancer cell lines

The cell line used for testing is the human colon cancer cell line HCT116. Cells are cultured in RPMI-1640 with 10-15% fetal calf serum at a concentration of 10,000 cells per milliliter in a volume of 0.5 ml and kept at 37 °C in a 95% air/5%CO_2 atmosphere.

Phosphorothioate oligoribonucleotides are synthesized on an Applied Biosystems Model 380B DNA synthesizer using phosphoroamidite chemistry. A sequence of 24 bases complementary to the nucleotides at position 1 to 24 of SEQ ID NO:1 is used as the test oligonucleotide. As a control, another (random) sequence is used: 5'-TCA ACT GAC TAG ATG TAC ATG GAC-3'. Following assembly and deprotection, oligonucleotides are ethanol-precipitated twice, dried, and suspended in phosphate buffered saline at the desired concentration. Purity of the oligonucleotides is tested by capillary gel electrophoresis and ion exchange HPLC. The purified oligonucleotides are added to the culture medium at a concentration of 10 µM once per day for seven days.

The addition of the test oligonucleotide for seven days results in significantly reduced expression of human serine palmitoyltransferase as determined by Western blotting. This effect is not observed with the control oligonucleotide. After 3 to 7 days, the number of cells in the cultures is counted using an automatic cell counter. The number of cells in cultures treated with the test oligonucleotide (expressed as 100%) is compared with the number of cells in cultures treated with the control oligonucleotide. The number of cells in cultures treated with the test oligonucleotide is not more than 30% of control, indicating that the inhibition of human serine palmitoyltransferase has an anti-proliferative effect on cancer cells.

EXAMPLE 7

In vivo testing of compounds/target validation

1. Acute Mechanistic Assays

1.1. Reduction in Mitogenic Plasma Hormone Levels

This non-tumor assay measures the ability of a compound to reduce either the endogenous level of a circulating hormone or the level of hormone produced in response to a biologic stimulus. Rodents are administered test compound (p.o., i.p., i.v., i.m., or s.c.). At a predetermined time after administration of test compound, blood plasma is collected. Plasma is assayed for levels of the hormone of interest. If the normal circulating levels of the hormone are too low and/or variable to provide consistent results, the level of the hormone may be elevated by a pre-treatment with a biologic stimulus (i.e., LHRH may be injected i.m. into mice at a dosage of 30 ng/mouse to induce a burst of testosterone synthesis). The timing of plasma collection would be adjusted to coincide with the peak of the induced hormone response. Compound effects are compared to a vehicle-treated control group. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test. Significance is p value ≤ 0.05 compared to the vehicle control group.

1.2. Hollow Fiber Mechanism of Action Assay

Hollow fibers are prepared with desired cell line(s) and implanted intraperitoneally and/or subcutaneously in rodents. Compounds are administered p.o., i.p., i.v., i.m., or s.c. Fibers are harvested in accordance with specific readout assay protocol, these may include assays for gene expression (bDNA, PCR, or Taqman), or a specific biochemical activity (i.e., cAMP levels. Results are analyzed by Student's t-test or Rank Sum test after the variance between groups is compared by an F-test, with significance at $p \leq 0.05$ as compared to the vehicle control group.

2. Subacute Functional *In Vivo* Assays

2.1. Reduction in Mass of Hormone Dependent Tissues

This is another non-tumor assay that measures the ability of a compound to reduce the mass of a hormone dependent tissue (i.e., seminal vesicles in males and uteri in females). Rodents are administered test compound (p.o., i.p., i.v., i.m., or s.c.) according to a predetermined schedule and for a predetermined duration (i.e., 1 week). At termination of the study, animals are weighed, the target organ is excised, any fluid is expressed, and the weight of the organ is recorded. Blood plasma may also be collected. Plasma may be assayed for levels of a hormone of interest or for levels of test agent. Organ weights may be directly compared or they may be normalized for the body weight of the animal. Compound effects are compared to a vehicle-treated control group. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test. Significance is p value ≤ 0.05 compared to the vehicle control group.

2.2. Hollow Fiber Proliferation Assay

Hollow fibers are prepared with desired cell line(s) and implanted intraperitoneally and/or subcutaneously in rodents. Compounds are administered p.o., i.p., i.v., i.m., or s.c. Fibers are harvested in accordance with specific readout assay protocol. Cell proliferation is determined by measuring a marker of cell number (i.e., MTT or LDH). The cell number and change in cell number from the starting inoculum are analyzed by Student's t-test or Rank Sum test after the variance between groups is compared by an F-test, with significance at $p \leq 0.05$ as compared to the vehicle control group.

2.3. Anti-angiogenesis Models

2.3.1. Corneal Angiogenesis

Hydron pellets with or without growth factors or cells are implanted into a micropocket surgically created in the rodent cornea. Compound administration may be systemic or local (compound mixed with growth factors in the hydron pellet). Corneas are harvested at 7 days post implantation immediately following intracardiac infusion of colloidal carbon and are fixed in 10% formalin. Readout is qualitative scoring and/or image analysis. Qualitative scores are compared by Rank Sum test. Image analysis data is evaluated by measuring the area of neovascularization (in pixels) and group averages are compared by Student's t-test (2 tail). Significance is $p \leq 0.05$ as compared to the growth factor or cells only group.

2.3.2. Matrigel Angiogenesis

Matrigel, containing cells or growth factors, is injected subcutaneously. Compounds are administered p.o., i.p., i.v., i.m., or s.c. Matrigel plugs are harvested at predetermined time point(s) and prepared for readout. Readout is an ELISA-based assay for hemoglobin concentration and/or histological examination (i.e. vessel count, special staining for endothelial surface markers: CD31, factor-8). Readouts are analyzed by Student's t-test, after the variance between groups is compared by an F-test, with significance determined at $p \leq 0.05$ as compared to the vehicle control group.

3. Primary Antitumor Efficacy

3.1. Early Therapy Models

3.1.1. Subcutaneous Tumor

Tumor cells or fragments are implanted subcutaneously on Day 0. Vehicle and/or compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule starting at a time, usually on Day 1, prior to the ability to measure the tumor

burden. Body weights and tumor measurements are recorded 2-3 times weekly. Mean net body and tumor weights are calculated for each data collection day. Anti-tumor efficacy may be initially determined by comparing the size of treated (T) and control (C) tumors on a given day by a Student's t-test, after the variance between groups is compared by an F-test, with significance determined at $p \leq 0.05$. The experiment may also be continued past the end of dosing in which case tumor measurements would continue to be recorded to monitor tumor growth delay. Tumor growth delays are expressed as the difference in the median time for the treated and control groups to attain a predetermined size divided by the median time for the control group to attain that size. Growth delays are compared by generating Kaplan-Meier curves from the times for individual tumors to attain the evaluation size. Significance is $p \leq 0.05$.

3.1.2. Intraperitoneal/Intracranial Tumor Models

Tumor cells are injected intraperitoneally or intracranially on Day 0. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule starting on Day 1. Observations of morbidity and/or mortality are recorded twice daily. Body weights are measured and recorded twice weekly. Morbidity/mortality data is expressed in terms of the median time of survival and the number of long-term survivors is indicated separately. Survival times are used to generate Kaplan-Meier curves. Significance is $p \leq 0.05$ by a log-rank test compared to the control group in the experiment.

3.2. Established Disease Model

Tumor cells or fragments are implanted subcutaneously and grown to the desired size for treatment to begin. Once at the predetermined size range, mice are randomized into treatment groups. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule. Tumor and body weights are measured and recorded 2-3 times weekly. Mean tumor weights of all groups over days post inocu-

lation are graphed for comparison. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test to compare tumor sizes in the treated and control groups at the end of treatment. Significance is $p \leq 0.05$ as compared to the control group. Tumor measurements may be recorded after dosing has stopped to monitor tumor growth delay. Tumor growth delays are expressed as the difference in the median time for the treated and control groups to attain a predetermined size divided by the median time for the control group to attain that size. Growth delays are compared by generating Kaplan-Meier curves from the times for individual tumors to attain the evaluation size. Significance is p value≤ 0.05 compared to the vehicle control group.

3.3. Orthotopic Disease Models

3.3.1. Mammary Fat Pad Assay

Tumor cells or fragments, of mammary adenocarcinoma origin, are implanted directly into a surgically exposed and reflected mammary fat pad in rodents. The fat pad is placed back in its original position and the surgical site is closed. Hormones may also be administered to the rodents to support the growth of the tumors. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule. Tumor and body weights are measured and recorded 2-3 times weekly. Mean tumor weights of all groups over days post inoculation are graphed for comparison. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test to compare tumor sizes in the treated and control groups at the end of treatment. Significance is $p \leq 0.05$ as compared to the control group.

Tumor measurements may be recorded after dosing has stopped to monitor tumor growth delay. Tumor growth delays are expressed as the difference in the median time for the treated and control groups to attain a predetermined size divided by the median time for the control group to attain that size. Growth delays are compared by generating Kaplan-Meier curves from the times for individual tumors to attain the

evaluation size. Significance is p value≤ 0.05 compared to the vehicle control group. In addition, this model provides an opportunity to increase the rate of spontaneous metastasis of this type of tumor. Metastasis can be assessed at termination of the study by counting the number of visible foci per target organ, or measuring the target organ weight. The means of these endpoints are compared by Student's t-test after conducting an F-test, with significance determined at $p \leq 0.05$ compared to the control group in the experiment.

3.3.2. Intraprostatic Assay

Tumor cells or fragments, of prostatic adenocarcinoma origin, are implanted directly into a surgically exposed dorsal lobe of the prostate in rodents. The prostate is externalized through an abdominal incision so that the tumor can be implanted specifically in the dorsal lobe while verifying that the implant does not enter the seminal vesicles. The successfully inoculated prostate is replaced in the abdomen and the incisions throught e abdomen and skin are closed. Hormones may also be administered to the rodents to support the growth of the tumors. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule. Body weights are measured and recorded 2-3 times weekly. At a predetermined time, the experiment is terminated and the animal is dissected. The size of the primary tumor is measured in three dimensions using either a caliper or an ocular micrometer attached to a dissecting scope. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test to compare tumor sizes in the treated and control groups at the end of treatment. Significance is $p \leq 0.05$ as compared to the control group. This model provides an opportunity to increase the rate of spontaneous metastasis of this type of tumor. Metastasis can be assessed at termination of the study by counting the number of visible foci per target organ (i.e., the lungs), or measuring the target organ weight (i.e., the regional lymph nodes). The means of these endpoints are compared by Student's t-test after conducting an F-test, with significance determined at $p \leq 0.05$ compared to the control group in the experiment.

3.3.3. Intrabronchial Assay

Tumor cells of pulmonary origin may be implanted intrabronchially by making an incision through the skin and exposing the trachea. The trachea is pierced with the beveled end of a 25 gauge needle and the tumor cells are inoculated into the main bronchus using a flat-ended 27 gauge needle with a 90° bend. Compounds are ad-

ministered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule. Body weights are measured and recorded 2-3 times weekly. At a predetermined time, the experiment is terminated and the animal is dissected. The size of the primary tumor is measured in three dimensions using either a caliper or an ocular micrometer attached to a dissecting scope. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test to compare tumor sizes in the treated and control groups at the end of treatment. Significance is $p \leq 0.05$ as compared to the control group. This model provides an opportunity to increase the rate of spontaneous metastasis of this type of tumor. Metastasis can be assessed at termination of the study by counting the number of visible foci per target organ (i.e., the contralateral lung), or measuring the target organ weight. The means of these endpoints are compared by Student's t-test after conducting an F-test, with significance determined at $p \leq 0.05$ compared to the control group in the experiment.

3.3.4. Intracecal Assay

Tumor cells of gastrointestinal origin may be implanted intracecally by making an abdominal incision through the skin and externalizing the intestine. Tumor cells are inoculated into the cecal wall without penetrating the lumen of the intestine using a 27 or 30 gauge needle. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule. Body weights are measured and recorded 2-3 times weekly. At a predetermined time, the experiment is terminated and the animal is dissected. The size of the primary tumor is measured in three dimensions using either a caliper or an ocular micrometer attached to a dissecting scope. An F-test is preformed to determine if the variance is equal or unequal followed by a Student's t-test to compare tumor sizes in the treated and control groups at the end of treatment. Significance is $p \leq 0.05$ as compared to the control group. This model provides an opportunity to increase the rate of spontaneous metastasis of this type of tumor. Metastasis can be assessed at termination of the study by counting the number of visible foci per target organ (i.e., the liver), or measuring the target organ weight. The means of these endpoints are compared by Student's t-test after conducting an F-test,

with significance determined at $p \leq 0.05$ compared to the control group in the experiment.

4. Secondary (Metastatic) Antitumor Efficacy

4.1. Spontaneous Metastasis

Tumor cells are inoculated s.c. and the tumors allowed to grow to a predetermined range for spontaneous metastasis studies to the lung or liver. These primary tumors are then excised. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule which may include the period leading up to the excision of the primary tumor to evaluate therapies directed at inhibiting the early stages of tumor metastasis. Observations of morbidity and/or mortality are recorded daily. Body weights are measured and recorded twice weekly. Potential endpoints include survival time, numbers of visible foci per target organ, or target organ weight. When survival time is used as the endpoint the other values are not determined. Survival data is used to generate Kaplan-Meier curves. Significance is $p \leq 0.05$ by a log-rank test compared to the control group in the experiment. The mean number of visible tumor foci, as determined under a dissecting microscope, and the mean target organ weights are compared by Student's t-test after conducting an F-test, with significance determined at $p \leq 0.05$ compared to the control group in the experiment for both of these endpoints.

4.2. Forced Metastasis

Tumor cells are injected into the tail vein, portal vein, or the left ventricle of the heart in experimental (forced) lung, liver, and bone metastasis studies, respectively. Compounds are administered p.o., i.p., i.v., i.m., or s.c. according to a predetermined schedule. Observations of morbidity and/or mortality are recorded daily. Body weights are measured and recorded twice weekly. Potential endpoints include survival time, numbers of visible foci per target organ, or target organ weight. When

survival time is used as the endpoint the other values are not determined. Survival data is used to generate Kaplan-Meier curves. Significance is $p \leq 0.05$ by a log-rank test compared to the control group in the experiment. The mean number of visible tumor foci, as determined under a dissecting microscope, and the mean target organ weights are compared by Student's t-test after conducting an F-test, with significance at $p \leq 0.05$ compared to the vehicle control group in the experiment for both end-points.

EXAMPLE 8

In vivo testing of compounds/target validation

1. Pain:

Acute Pain

Acute pain is measured on a hot plate mainly in rats. Two variants of hot plate testing are used: In the classical variant animals are put on a hot surface (52 to 56 C) and the latency time is measured until the animals show nocifensive behavior, such as stepping or foot licking. The other variant is an increasing temperature hot plate where the experimental animals are put on a surface of neutral temperature. Subsequently this surface is slowly but constantly heated until the animals begin to lick a hind paw. The temperature which is reached when hind paw licking begins is a measure for pain threshold.

Compounds are tested against a vehicle treated control group. Substance application is performed at different time points via different application routes (i.v., i.p., p.o., i.t., i.c.v., s.c., intradermal, transdermal) prior to pain testing.

Persistent Pain

Persistent pain is measured with the formalin or capsaicin test, mainly in rats. A solution of 1 to 5% formalin or 10 to 100 µg capsaicin is injected into one hind paw of the experimental animal. After formalin or capsaicin application the animals show

nocifensive reactions like flinching, licking and biting of the affected paw. The number of nocifensive reactions within a time frame of up to 90 minutes is a measure for intensity of pain.

Compounds are tested against a vehicle treated control group. Substance application is performed at different time points via different application routes (i.v., i.p., p.o., i.t., i.c.v., s.c., intradermal, transdermal) prior to formalin or capsaicin administration.

Neuropathic Pain

Neuropathic pain is induced by different variants of unilateral sciatic nerve injury mainly in rats. The operation is performed under anesthesia. The first variant of sciatic nerve injury is produced by placing loosely constrictive ligatures around the common sciatic nerve. The second variant is the tight ligation of about the half of the diameter of the common sciatic nerve. In the next variant, a group of models is used in which tight ligations or transections are made of either the L5 and L6 spinal nerves, or the L% spinal nerve only. The fourth variant involves an axotomy of two of the three terminal branches of the sciatic nerve (tibial and common peroneal nerves) leaving the remaining sural nerve intact whereas the last variant comprises the axotomy of only the tibial branch leaving the sural and common nerves uninjured. Control animals are treated with a sham operation.

Postoperatively, the nerve injured animals develop a chronic mechanical allodynia, cold allodynioa, as well as a thermal hyperalgesia. Mechanical allodynia is measured by means of a pressure transducer (electronic von Frey Anesthesiometer, IITC Inc.-Life Science Instruments, Woodland Hills, SA, USA; Electronic von Frey System, Somedic Sales AB, Hörby, Sweden). Thermal hyperalgesia is measured by means of a radiant heat source (Plantar Test, Ugo Basile, Comerio, Italy), or by means of a cold plate of 5 to 10 °C where the nocifensive reactions of the affected hind paw are counted as a measure of pain intensity. A further test for cold induced pain is the counting of nocifensive reactions, or duration of nocifensive responses

after plantar administration of acetone to the affected hind limb. Chronic pain in general is assessed by registering the circadanian rhytms in activity (Surjo and Arndt, Universität zu Köln, Cologne, Germany), and by scoring differences in gait (foot print patterns; FOOTPRINTS program, Klapdor et al., 1997. A low cost method to analyse footprint patterns. J. Neurosci. Methods 75, 49-54).

Compounds are tested against sham operated and vehicle treated control groups. Substance application is performed at different time points via different application routes (i.v., i.p., p.o., i.t., i.c.v., s.c., intradermal, transdermal) prior to pain testing.

Inflammatory Pain

Inflammatory pain is induced mainly in rats by injection of 0.75 mg carrageenan or complete Freund's adjuvant into one hind paw. The animals develop an edema with mechanical allodynia as well as thermal hyperalgesia. Mechanical allodynia is measured by means of a pressure transducer (electronic von Frey Anesthesiometer, ITC Inc.-Life Science Instruments, Woodland Hills, SA, USA). Thermal hyperalgesia is measured by means of a radiant heat source (Plantar Test, Ugo Basile, Comerio, Italy, Paw thermal stimulator, G. Ozaki, University of California, USA). For edema measurement two methods are being used. In the first method, the animals are sacrificed and the affected hindpaws sectioned and weighed. The second method comprises differences in paw volume by measuring water displacement in a plethysmometer (Ugo Basile, Comerio, Italy).

Compounds are tested against uninflamed as well as vehicle treated control groups. Substance application is performed at different time points via different application routes (i.v., i.p., p.o., i.t., i.c.v., s.c., intradermal, transdermal) prior to pain testing.

Diabetic Neuropathic Pain

Rats treated with a single intraperitoneal injection of 50 to 80 mg/kg streptozotocin develop a profound hyperglycemia and mechanical allodynia within 1 to 3 weeks. Mechanical allodynia is measured by means of a pressure transducer (electronic von Frey Anesthesiometer, IITC Inc.-Life Science Instruments, Woodland Hills, SA, USA).

Compounds are tested against diabetic and non-diabetic vehicle treated control groups. Substance application is performed at different time points via different application routes (i.v., i.p., p.o., i.t., i.c.v., s.c., intradermal, transdermal) prior to pain testing.

2. Parkinson's disease

6-Hydroxydopamine (6-OH-DA) Lesion

Degeneration of the dopaminergic nigrostriatal and striatopallidal pathways is the central pathological event in Parkinson's disease. This disorder has been mimicked experimentally in rats using single/sequential unilateral stereotaxic injections of 6-OH-DA into the medium forebrain bundle (MFB).

Male Wistar rats (Harlan Winkelmann, Germany), weighing 200±250 g at the beginning of the experiment, are used. The rats are maintained in a temperature- and humidity-controlled environment under a 12 h light/dark cycle with free access to food and water when not in experimental sessions. The following in vivo protocols are approved by the governmental authorities. All efforts are made to minimize animal suffering, to reduce the number of animals used, and to utilize alternatives to in vivo techniques.

Animals are administered pargyline on the day of surgery (Sigma, St. Louis, MO, USA; 50 mg/kg i.p.) in order to inhibit metabolism of 6-OHDA by monoamine oxi-

dase and desmethylimipramine HCl (Sigma; 25 mg/kg i.p.) in order to prevent uptake of 6-OHDA by noradrenergic terminals. Thirty minutes later the rats are anesthetized with sodium pentobarbital (50 mg/kg) and placed in a stereotaxic frame. In order to lesion the DA nigrostriatal pathway 4 µl of 0.01% ascorbic acid-saline containing 8 µg of 6-OHDA HBr (Sigma) are injected into the left medial fore-brain bundle at a rate of 1 µl/min (2.4 mm anterior, 1.49 mm lateral, -2.7 mm ventral to Bregma and the skull surface). The needle is left in place an additional 5 min to allow diffusion to occur.

Stepping Test

Forelimb akinesia is assessed three weeks following lesion placement using a modified stepping test protocol. In brief, the animals are held by the experimenter with one hand fixing the hindlimbs and slightly raising the hind part above the surface. One paw is touching the table, and is then moved slowly sideways (5 s for 1 m), first in the forehand and then in the backhand direction. The number of adjusting steps is counted for both paws in the backhand and forehand direction of movement. The sequence of testing is right paw forehand and backhand adjusting stepping, followed by left paw forehand and backhand directions. The test is repeated three times on three consecutive days, after an initial training period of three days prior to the first testing. Forehand adjusted stepping reveals no consistent differences between lesioned and healthy control animals. Analysis is therefore restricted to backhand adjusted stepping.

Balance Test

Balance adjustments following postural challenge are also measured during the stepping test sessions. The rats are held in the same position as described in the stepping test and, instead of being moved sideways, tilted by the experimenter towards the side of the paw touching the table. This manoeuvre results in loss of balance and the ability of the rats to regain balance by forelimb movements is scored on a scale

ranging from 0 to 3. Score 0 is given for a normal forelimb placement. When the forelimb movement is delayed but recovery of postural balance detected, score 1 is given. Score 2 represents a clear, yet insufficient, forelimb reaction, as evidenced by muscle contraction, but lack of success in recovering balance, and score 3 is given for no reaction of movement. The test is repeated three times a day on each side for three consecutive days after an initial training period of three days prior to the first testing.

Staircase Test (Paw Reaching)

A modified version of the staircase test is used for evaluation of paw reaching behaviour three weeks following primary and secondary lesion placement. Plexiglass test boxes with a central platform and a removable staircase on each side are used. The apparatus is designed such that only the paw on the same side at each staircase can be used, thus providing a measure of independent forelimb use. For each test the animals are left in the test boxes for 15 min. The double staircase is filled with 7 x 3 chow pellets (Precision food pellets, formula: P, purified rodent diet, size 45 mg; Sandown Scientific) on each side. After each test the number of pellets eaten (successfully retrieved pellets) and the number of pellets taken (touched but dropped) for each paw and the success rate (pellets eaten/pellets taken) are counted separately. After three days of food deprivation (12 g per animal per day) the animals are tested for 11 days. Full analysis is conducted only for the last five days.

MPTP treatment

The neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydro-pyridine (MPTP) causes degeneration of mesencephalic dopaminergic (DAergic) neurons in rodents, non-human primates, and humans and, in so doing, reproduces many of the symptoms of Parkinson's disease. MPTP leads to a marked decrease in the levels of dopamine and its metabolites, and in the number of dopaminergic terminals in the striatum as well as severe loss of the tyrosine hydroxylase (TH)-immunoreactive cell bodies in the substantia nigra, pars compacta.

In order to obtain severe and long-lasting lesions, and to reduce mortality, animals receive single injections of MPTP, and are then tested for severity of lesion 7–10 days later. Successive MPTP injections are administered on days 1, 2 and 3. Animals receive application of 4 mg/kg MPTP hydrochloride (Sigma) in saline once daily. All injections are intraperitoneal (i.p.) and the MPTP stock solution is frozen between injections. Animals are decapitated on day 11.

Immunohistology

At the completion of behavioral experiments, all animals are anaesthetized with 3 ml thiopental (1 g/40 ml i.p., Tyrol Pharma). The mice are perfused transcardially with 0.01 M PBS (pH 7.4) for 2 min, followed by 4% paraformaldehyde (Merck) in PBS for 15 min. The brains are removed and placed in 4% paraformaldehyde for 24 h at 4°C. For dehydration they are then transferred to a 20% sucrose (Merck) solution in 0.1 M PBS at 4 °C until they sink. The brains are frozen in methylbutan at -20 °C for 2 min and stored at -70 °C. Using a sledge microtome (mod. 3800-Frigocut, Leica), 25 µm sections are taken from the genu of the corpus callosum (AP 1.7 mm) to the hippocampus (AP 21.8 mm) and from AP 24.16 to AP 26.72. Forty-six sections are cut and stored in assorters in 0.25 M Tris buffer (pH 7.4) for immunohistochemistry.

A series of sections is processed for free-floating tyrosine hydroxylase (TH) immunohistochemistry. Following three rinses in 0.1 M PBS, endogenous peroxidase activity is quenched for 10 min in 0.3% H_2O_2 ±PBS. After rinsing in PBS, sections are preincubated in 10% normal bovine serum (Sigma) for 5 min as blocking agent and transferred to either primary anti-rat TH rabbit antiserum (dilution 1:2000).

Following overnight incubation at room temperature, sections for TH mmunoreactivity are rinsed in PBS (2 x10 min) and incubated in biotinylated anti-rabbit immunoglobulin G raised in goat (dilution 1:200) (Vector) for 90 min, rinsed repeatedly and transferred to Vectastain ABC (Vector) solution for 1 h. 3,.3'

-Diaminobenzidine tetrahydrochloride (DAB; Sigma) in 0.1 M PBS, supplemented with 0.005% H_2O_2 , serves as chromogen in the subsequent visualization reaction. Sections are mounted on to gelatin-coated slides, left to dry overnight, counter-stained with hematoxylin dehydrated in ascending alcohol concentrations and cleared in butylacetate. Coverslips are mounted on entellan.

Rotarod Test

We use a modification of the procedure described by Rozas and Labandeira-Garcia (1997), with a CR-1 Rotamex system (Columbus Instruments, Columbus, OH) comprising an IBM-compatible personal computer, a CIO-24 data acquisition card, a control unit, and a four-lane rotarod unit. The rotarod unit consists of a rotating spindle (diameter 7.3 cm) and individual compartments for each mouse. The system software allows preprogramming of session protocols with varying rotational speeds (0–80 rpm). Infrared beams are used to detect when a mouse has fallen onto the base grid beneath the rotarod. The system logs the fall as the end of the experiment for that mouse, and the total time on the rotarod, as well as the time of the fall and all the set-up parameters, are recorded. The system also allows a weak current to be passed through the base grid, to aid training.

3. Dementia

The object recognition task

The object recognition task has been designed to assess the effects of experimental manipulations on the cognitive performance of rodents. A rat is placed in an open field, in which two identical objects are present. The rats inspects both objects during the first trial of the object recognition task. In a second trial, after a retention interval of for example 24 hours, one of the two objects used int the first trial, the 'familiar' object, and a novel object are placed in the open field. The inspection time at each of the objects is registered. The basic measures in the OR task is the time spent by a rat

exploring the two object the second trial. Good retention is reflected by higher explortation times towards the novel than the 'familiar' object.

Administration of the putative cognition enhancer prior to the first trial predominantly allows assessment of the effects on acquisition, and eventually on consolidation processes. Administration of the testing compound after the first trial allows to assess the effects on consolidation processes, whereas administration before the second trial allows to measure effects on retrieval processes.

The passive avoidance task

The passive avoidance task assesses memory performance in rats and mice. The inhibitory avoidance apparatus consists of a two-compartment box with a light compartment and a dark compartment. The two compartments are separated by a guillotine door that can be operated by the experimenter. A threshold of 2 cm separates the two compartments when the guillotine door is raised. When the door is open, the illumination in the dark compartment is about 2 lux. The light intensity is about 500 lux at the center of the floor of the light compartment.

Two habituation sessions, one shock session, and a retention session are given, separated by inter-session intervals of 24 hours. In the habituation sessions and the retention session the rat is allowed to explore the apparatus for 300 sec. The rat is placed in the light compartment, facing the wall opposite to the guillotine door. After an accommodation period of 15 sec. the guillotine door is opened so that all parts of the apparatus can be visited freely. Rats normally avoid brighly lit areas and will enter the dark compartment within a few seconds.

In the shock session the guillotine door between the compartments is lowered as soon as the rat has entered the dark compartment with its four paws, and a scrambled 1 mA footshock is administered for 2 sec. The rat is removed from the apparatus and

put back into its home cage. The procedure during the retention session is identical to that of the habituation sessions.

The step-through latency, that is the first latency of entering the dark compartment (in sec.) during the retention session is an index of the memory performance of the animal; the longer the latency to enter the dark compartment, the better the retention is. A testing compound in given half an hour before the shock session, together with 1 mg*kg^{-1} scopolamine. Scopolamine impairs the memory performance during the retention session 24 hours later. If the test compound increases the enter latency compared with the scopolamine-treated controls, is is likely to possess cognition enhancing potential.

The Morris water escape task

The Morris water escape task measures spatial orientation learning in rodents. It is a test system that has extensively been used to investigate the effects of putative therapeutic on the cognitive functions of rats and mice. The performance of an animal is assessed in a circular water tank with an escape platform that is submerged about 1 cm below the surface of the water. The escape platform is not visible for an animal swimming in the water tank. Abundant extra-maze cues are provided by the furniture in the room, including desks, computer equipment, a second water tank, the presence of the experimenter, and by a radio on a shelf that is playing softly.

The animals receive four trials during five daily acquisition sessions. A trial is started by placing an anmimal into the pool, facing the wall of the tank. Each of four starting positions in the quadrants north, east, south, and west is used once in a series of four trials; their order is randomized. The escape platform is always in the same position. A trial is terminated as soon as the animal had climbs onto the escape platform or when 90 seconds have elapsed, whichever event occurs first. Teh animal is allowed to stay on the platform for 30 seconds. Then it is taken from the platform and the next trial is started. If an amimal did not find the platform within 90 seconds it is put

on the platform by the experimenter and is allowed to stay there for 30 seconds. After the fourth trial of the fifth daily session, an additional trial is given as a probe trial: the platform is removed, and the time the animal spents in the four quadrants is measured for 30 or 60 seconds. In the probe trial, all animals start from the same start position, opposite to the quadrant where the escape platform had been positioned during acquisition.

Four different measures are taken to evaluate the performance of an animal during acquisition training: escape latency, traveled distance, distance to platform, and swimming speed. The following measures are evaluated for the probe trial: time (s) in quadrants and traveled distance (cm) in the four quadrants. The probe trial provides additional information about how well an animal learned the position of the escape platform. If an animal spents more time and swims a longer distance in the quadrant where the platform had been positioned during the acquisition sessions than in any other quadrant, one concludes that the platform position has been learned well.

In order to assess the effects of putative congition enhacing compounds, rats or mice with specific brain lesions which impair cognitive functions, or animals treated with compounds such as scopolamine or MK-801, which interfere with normal learning, or aged animals which suffer from cognitive deficits, are used.

The T-maze spontaneous alternation task

The T-maze spontaneous alternation task (TeMCAT) assesses the spatial memory performance in mice. The start arm and the two goal arms of the T-maze are provided with guillotine doors which can be operated manually by the experimenter. A mouse is put into the start arm at the beginning of training. The guillotine door is closed. In the first trial, the 'forced trial', either the left or right goal arm is blocked by lowering the guillotine door. After the mouse has been released from the start arm, it will negotiate the maze, eventually enter the open goal arm, and return to the start position, where it will be confined for 5 seconds, by lowering the guillotine

door. Then, the animal can choose freely between the left and right goal arm (all guillotine-doors opened) diring 14 'free choice' trials. As soon a the mouse has entered one goal arm, the other one is closed. The mouse eventually returns to the start arm and is free to visit whichever goalarm it wants after having been confined to the start arm for 5 seconds. After completion of 14 free choice trials in one session, the animal is removed from the maze. During training, the animal is never handeled.

The per-cent alternations out of 14 trials is calculated. This per-centage and the total time needed to complete the first forced trial and the subsequent 14 free choice trials (in s) is analysed. Cognitive deficits are usually induced by an injection of scopolamine, 30 min before the start of the training session. Scopolamine reduced the per-cent alternations to chance level, or below. A cognition enhancer, which is always administered before the training session, will at least partially, antagonize the scopolamine-induced reduction in the spontaneous alternation rate.

EXAMPLE 9

Tissue-specific expression of matriptase-like serine protease

To to establish a role for matriptase-like serine protease in the pathogenesis of COPD, expression profiling of the gene was done using real-time quantitative PCR with RNA samples from human respiratory tissues and inflammatory cells relevant to COPD. The panel consisted of total RNA samples lung (adult and fetal), trachea, freshly isolated alveolar type II cells, cultured human bronchial epithelial cells, cultured small airway epithelial cells, cultured bronchial sooth muscle cells, cultured H441 cells (Clara-like), freshly isolated neutrophils and monocytes, and cultured monocytes (macrophage-like). Expression of matriptase-like serine protease also was evaluated in a range of human tissues using total RNA panels obtained from Clontech Laboratories, UK, Ltd.. The tissues were adrenal gland, bone marrow, brain, colon, heart, kidney, liver, lung, mammary gland, pancreas, prostate, salivary gland, skeletal muscle, small intesting, spleen, stomach, testis, thymus, trachea, thyroid, and uterus.

Real-time quantitative PCR. Expression profiling of the target gene was performed using real-time quantitative PCR, a development of the kinetic analysis of PCR first described in Higuchi *et al.*, *BioTechnology 10*, 413-17, 1992, and Higuchi *et al.*, *BioTechnology 11*, 1026-30, 1993. The principle is that at any given cycle within the exponential phase of PCR, the amount of product is proportional to the initial number of template copies.

PCR amplification is performed in the presence of an oligonucleotide probe (TaqMan probe) that is complementary to the target sequence and labeled with a fluorescent reporter dye and a quencher dye. During the extension phase of PCR, the probe is cleaved by the 5'-3' endonuclease activity of Taq DNA polymerase, releasing the fluorophore from the effect of the quenching dye (Holland *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 88*, 7276-80, 1991). Because the fluorescence emission increases in direct proportion to the amount of the specific amplified product, the exponential growth phase of PCR product can be detected and used to determine the initial template concentration (Heid *et al.*, *Genome Res. 6*, 986-94, 1996, and Gibson *et al.*, *Genome Res. 6*, 995-1001, 1996).

Real-time quantitative PCR was done using an ABI Prism 7700 Sequence Detector. The C_T value generated for each reaciton was used to determine the initial template concentration (copy number) by interpolation from a universal standard curve. The level of expression of the target gene in each sample was calculated relative to the sample with the lowest expression of the gene.

RNA extraction and cDNA preparation. Total RNA from each of the respiratory tissues and inflammatory cell types listed above were isolated using Qiagen's RNeasy system according to the manufacturer's protocol (Crawley, West Sussex, UK). The concentration of purified RNA was determined using a RiboGreen RNA quantitation kit (Molecular Probes Europe, The Netherlands). For the preparation of cDNA, 1 µg of total RNA was reverse transcribed in a final volume of 20 µl, using

200 U of SUPERSCRIPT™ RNase H⁻ Reverse Transcriptase (Life Technologies, Paisley, UK), 10 mM dithiothreitol, 0.5 mM of each dNTP and 5 μM random hexamers (Applied Biosystems, Warrington, Cheshire, UK) according to the manufacturer's protocol.

TaqMan quantitative analysis. Specific primers and probe were designed according to the recommendations of PE Applied Biosystems. The probe was labeled at the 5' end with FAM (6-carboxyfluorescein). Quantification PCR was performed with 5 ng of reverse transcribed RNA from each sample. Each determination is done in duplicate.

The assay reaction mix was as follows: 1X final TaqMan Universal PCR Master Mix (from 2X stock) (PE Applied Biosystems, CA); 900 nM forward primer; 900 nM reverse primer; 200 nM probe; 5 ng cDNA; and water to 25 μl.

Each of the following steps were carried out once: pre PCR, 2 minutes at 50° C, and 10 minutes at 95°C. The following steps are carried out 40 times: denaturation, 15 seconds at 95°C, annealing/extension, 1 minute at 60°C.

All experiments were performed using an ABI Prism 7700 Sequence Detector (PE Applied Biosystems, CA). At the end of the run, fluorescence data acquired during PCR were processed as described in the ABI Prism 7700 user's manual to achieve better background subtraction as well as signal linearity with the starting target quantity.

Tables 1 and 2 show the results of expression profiling for matriptase-like serine protease using the indicated cell and tissue samples. For Table 1, the cells are defined as follows: HBEC, cultured human bronchial epithelial cells; H441, a Clara-like cell line; SAE, cultured small airway epithelial cells; SMC, cultured airway smooth muscle cells; AII, freshly isolated human alveolar type II cells; Neut, freshly isolated circulating neutrophils; Mono, freshly isolated monocytes; and CM, cultured

monocytes. Other letters identify the donor. The results are shown graphically in Figs. 12 and 13.

Table 1.

Tissue	Relative expression
Lung	49.20042452
Trachea	61.45721134
HBEC 1	21.3988212
HBEC 2	18.37159446
H441	6.519927252
SMC	3.282061011
SAE	1.305950302
AII	2.729626652
Foetal lung	3.021798839
COPD Neut 1	1.187200443
COPD Neut 2	1.107037341
COPD Neut 4	1.079248466
GAP Neut	0
AEM Neut	2.299213231
AT Neut	0
KN Neut	1.339572891
SM Mono	0
DLF Mono	2.02477211
DS Mono	2.561540637
RLH CM	1.365361177
CTP CM	1

Table 2.

Tissue	Relative expression
Adrenal gland	8.196114181
Bone Marrow	3.750688527
Brain	19.08565196
Colon	102.8359584
Heart	1
HL60	0.765727951
Kidney	6.90373321
Liver	1.560308233
Lung	23.39004191
Mammary gland	15.2792487
Pancreas	1.281284286
Prostate	13.62758815
Salivary gland	22.23059056
Skeletal Muscle	15.9743428
Sm Intest	1.973946601
Spleen	23.53921632
Stomach	3.324045807
Testis	204.2871266
Thymus	26.05877524
Thyroid	34.03141444
Uterus	1.121198996

REFERENCES

1. Nicolson (1988) Organ specificity of tumor metastasis: Role of preferential adhesion, invasion and growth of malignant cells at specific secondary sites. *Cancer Met. Rev.* 7, 143-188.

2. Liotta *et al.* (1983) Tumor invasion and the extracellular matrix. *Lab. Invest. 49*, 639-649.

3. Vlodavsky *et al.* (1987) Endothelial cell-derived basic fibroblast growth factor: Synthesis and deposition into subendothelial extracellular matrix. *Proc. Natl. Acad. Sci. USA 84*, 2292-2296.

4. Folkman *et al.* (1980) A heparin-binding angiogenic protein--basic fibroblast growth factor--is stored within basement membrane. *Am. J. Pathol. 130, 393400.*

5. Cardon-Cardo *et al.* (1990) Expression of basic fibroblast growth factor in normal human tissues. *Lab. Invest. 63*, 832-840.

6. Vlodavsky *et al.* (1991) Extracellular sequestration and release of fibroblast growth factor: a regulatory mechanism? *Trends Biochem. Sci. 16*, 268-271.

7. Vlodavsky *et al.* (1993) Extracellular matrix-bound growth factors, enzymes and plasma proteins. In BASEMENT MEMBRANES: CELLULAR AND MOLECULAR ASPECTS Rohrbach & Timpl, eds., pp 327-343. Academic Press Inc., Orlando, Fla.

8. Ross (1993) The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature (Lond.). 362*, 801-809.

9. McCarthy *et al.* (1986). Human fibronectin contains distinct adhesion- and motility-promoting domains for metastatic melanoma cells. *J. Cell Biol. 102*, 179-88.

10. Van Muijen *et al.* (1995) Properties of metastasizing and non-metastasizing human melanoma cells. *Recent Results in Cancer Research 139*, 104-22.

11. Price *et al.* (1997) The Biochemistry of Cancer Dissemination, in *Critical Reviews in Biochemistry and Mol. Biol. 32,* 175-253.

CLAIMS

1. An isolated polynucleotide encoding a matriptase-like serine protease polypeptide and being selected from the group consisting of:

a) a polynucleotide encoding a matriptase-like serine protease polypeptide comprising an amino acid sequence selected form the group consisting of:
amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 2;
the amino acid sequence shown in SEQ ID NO: 2;
amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO: 18; and
the amino acid sequence shown in SEQ ID NO: 18.

b) a polynucleotide comprising the sequence of SEQ ID NOS: 1 or 17;

c) a polynucleotide which hybridizes under stringent conditions to a polynucleotide specified in (a) and (b);

d) a polynucleotide the sequence of which deviates from the polynucleotide sequences specified in (a) to (c) due to the degeneration of the genetic code; and

e) a polynucleotide which represents a fragment, derivative or allelic variation of a polynucleotide sequence specified in (a to (d):

2. An expression vector containing any polynucleotide of claim 1.

3. A host cell containing the expression vector of claim 2.

4. A substantially purified matriptase-like serine protease polypeptide encoded by a polynucleotide of claim 1.

5. A method for producing a matriptase-like serine protease polypeptide, wherein the method comprises the following steps:

a) culturing the host cell of claim 3 under conditions suitable for the expression of the matriptase-like serine protease polypeptide; and

b) recovering the matriptase-like serine protease polypeptide from the host cell culture.

6. A method for detection of a polynucleotide encoding a matriptase-like serine protease polypeptide in a biological sample comprising the following steps:

a) hybridizing any polynucleotide of claim 1 to a nucleic acid material of a biological sample, thereby forming a hybridization complex; and

b) detecting said hybridization complex.

7. The method of claim 6, wherein before hybridization, the nucleic acid material of the biological sample is amplified.

8. A method for the detection of a polynucleotide of claim 1 or a matriptase-like serine protease polypeptide of claim 4 comprising the steps of:
contacting a biological sample with a reagent which specifically interacts with the polynucleotide or the matriptase-like serine protease polypeptide.

9. A diagnostic kit for conducting the method of any one of claims 6 to 8.

10. A method of screening for agents which decrease the activity of a matriptase-like serine protease, comprising the steps of:
contacting a test compound with any matriptase-like serine protease polypeptide encoded by any polynucleotide of claim1;
detecting binding of the test compound to the matriptase-like serine protease polypeptide, wherein a test compound which binds to the polypeptide is identified as a potential therapeutic agent for decreasing the activity of a matriptase-like serine protease.

11. A method of screening for agents which regulate the activity of a matriptase-like serine protease, comprising the steps of:
contacting a test compound with a matriptase-like serine protease polypeptide encoded by any polynucleotide of claim 1; and
detecting a matriptase-like serine protease activity of the polypeptide, wherein a test compound which increases the matriptase-like serine protease activity is identified as a potential therapeutic agent for increasing the activity of the matriptase-like serine protease, and wherein a test compound which decreases the matriptase-like serine protease activity of the polypeptide is identified as a potential therapeutic agent for decreasing the activity of the matriptase-like serine protease.

12. A method of screening for agents which decrease the activity of a matriptase-like serine protease, comprising the steps of:
contacting a test compound with any polynucleotide of claim 1 and detecting binding of the test compound to the polynucleotide, wherein a test compound which binds to the polynucleotide is identified as a potential therapeutic agent for decreasing the activity of matriptase-like serine protease.

13. A method of reducing the activity of matriptase-like serine protease, comprising the steps of:
contacting a cell with a reagent which specifically binds to any polynucleotide of claim 1 or any matriptase-like serine protease polypeptide of claim 4; whereby the activity of matriptase-like serine protease is reduced.

14. A reagent that modulates the activity of a matriptase-like serine protease polypeptide or a polynucleotide wherein said reagent is identified by the method of any of the claim 10 to 12.

15. A pharmaceutical composition, comprising:

the expression vector of claim 2 or the reagent of claim 14 and a pharmaceutically acceptable carrier.

16. Use of the pharmaceutical composition of claim 15 for modulating the activity of a matriptase-like serine protease in a disease.

17. Use of claim 16 wherein the disease is cancer, chronic obstructive pulmonary disease, a peripheral or central nervous system disorder, or a cardiovascular disorder.

18. A cDNA encoding a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 and 18.

19. The cDNA of claim 18 which comprises SEQ ID NOS:1 or 17.

20. The cDNA of claim 18 which consists of SEQ ID NOS:1 or 17.

21. An expression vector comprising a polynucleotide which encodes a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18.

22. The expression vector of claim 21 wherein the polynucleotide consists of SEQ ID NOS:1 or 17.

23. A host cell comprising an expression vector which encodes a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18.

24. The host cell of claim 23 wherein the polynucleotide consists of SEQ ID NOS:1 or 17.

25. A purified polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18.

26. The purified polypeptide of claim 25 which consists of the amino acid sequence shown in SEQ ID NO:2 or 18.

27. A fusion protein comprising a polypeptide having the amino acid sequence shown in SEQ ID NO:2 or 18.

28. A method of producing a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18, comprising the steps of:
culturing a host cell comprising an expression vector which encodes the polypeptide under conditions whereby the polypeptide is expressed; and
isolating the polypeptide.

29. The method of claim 28 wherein the expression vector comprises SEQ ID NOS:1 or 17.

30. A method of detecting a coding sequence for a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18, comprising the steps of:
hybridizing a polynucleotide comprising 11 contiguous nucleotides of SEQ ID NOS:1 or 17 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and
detecting the hybridization complex.

31. The method of claim 30 further comprising the step of amplifying the nucleic acid material before the step of hybridizing.

32. A kit for detecting a coding sequence for a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18, comprising:
a polynucleotide comprising 11 contiguous nucleotides of SEQ ID NOS:1 or 17; and
instructions for the method of claim 30.

33. A method of detecting a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18, comprising the steps of:
contacting a biological sample with a reagent that specifically binds to the polypeptide to form a reagent-polypeptide complex; and
detecting the reagent-polypeptide complex.

34. The method of claim 33 wherein the reagent is an antibody.

35. A kit for detecting a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18, comprising:
an antibody which specifically binds to the polypeptide; and
instructions for the method of claim 33.

36. A method of screening for agents which can modulate the activity of a human matriptase-like serine protease, comprising the steps of:
contacting a test compound with a polypeptide comprising an amino acid sequence selected from the group consisting of: (1) amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO:2 or 18 and (2) the amino acid sequence shown in SEQ ID NO:2 or 18; and
detecting binding of the test compound to the polypeptide, wherein a test compound which binds to the polypeptide is identified as a potential agent for regulating activity of the human matriptase-like serine protease.

37. The method of claim 36 wherein the step of contacting is in a cell.

38. The method of claim 36 wherein the cell is *in vitro*.

39. The method of claim 36 wherein the step of contacting is in a cell-free system.

40. The method of claim 36 wherein the polypeptide comprises a detectable label.

41. The method of claim 36 wherein the test compound comprises a detectable label.

42. The method of claim 36 wherein the test compound displaces a labeled ligand which is bound to the polypeptide.

43. The method of claim 36 wherein the polypeptide is bound to a solid support.

44. The method of claim 36 wherein the test compound is bound to a solid support.

45. A method of screening for agents which modulate an activity of a human matriptase-like serine protease, comprising the steps of:
contacting a test compound with a polypeptide comprising an amino acid sequence selected from the group consisting of: (1) amino acid sequences which are at least about 50% identical to the amino acid sequence shown in SEQ ID NO:2 or 18 and (2) the amino acid sequence shown in SEQ ID NO:2 or 18; and
detecting an activity of the polypeptide, wherein a test compound which increases the activity of the polypeptide is identified as a potential agent for increasing the activity of the human matriptase-like serine protease, and wherein a test compound which decreases the activity of the polypeptide is identified as a potential agent for decreasing the activity of the human matriptase-like serine protease.

46. The method of claim 45 wherein the step of contacting is in a cell.

47. The method of claim 45 wherein the cell is *in vitro*.

48. The method of claim 45 wherein the step of contacting is in a cell-free system.

49. A method of screening for agents which modulate an activity of a human matriptase-like serine protease, comprising the steps of:
contacting a test compound with a product encoded by a polynucleotide which comprises the nucleotide sequence shown in SEQ ID NOS:1 or 17; and detecting binding of the test compound to the product, wherein a test compound which binds to the product is identified as a potential agent for regulating the activity of the human matriptase-like serine protease.

50. The method of claim 49 wherein the product is a polypeptide.

51. The method of claim 49 wherein the product is RNA.

52. A method of reducing activity of a human matriptase-like serine protease, comprising the step of:
contacting a cell with a reagent which specifically binds to a product encoded by a polynucleotide comprising the nucleotide sequence shown in SEQ ID NOS:1 or 17, whereby the activity of a human matriptase-like serine protease is reduced.

53. The method of claim 52 wherein the product is a polypeptide.

54. The method of claim 53 wherein the reagent is an antibody.

55. The method of claim 52 wherein the product is RNA.

56. The method of claim 55 wherein the reagent is an antisense oligonucleotide.

57. The method of claim 56 wherein the reagent is a ribozyme.

58. The method of claim 52 wherein the cell is *in vitro*.

59. The method of claim 52 wherein the cell is *in vivo*.

60. A pharmaceutical composition, comprising:
a reagent which specifically binds to a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18; and
a pharmaceutically acceptable carrier.

61. The pharmaceutical composition of claim 60 wherein the reagent is an antibody.

62. A pharmaceutical composition, comprising:
a reagent which specifically binds to a product of a polynucleotide comprising the nucleotide sequence shown in SEQ ID NOS:1 or 17; and
a pharmaceutically acceptable carrier.

63. The pharmaceutical composition of claim 62 wherein the reagent is a ribozyme.

64. The pharmaceutical composition of claim 62 wherein the reagent is an antisense oligonucleotide.

65. The pharmaceutical composition of claim 62 wherein the reagent is an antibody.

66. A pharmaceutical composition, comprising:
an expression vector encoding a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2 or 18; and

a pharmaceutically acceptable carrier.

67. The pharmaceutical composition of claim 66 wherein the expression vector comprises SEQ ID NOS:1 or 17.

68. A method of treating a matriptase-like serine protease dysfunction related disease, wherein the disease is selected from cancer, chronic obstructive pulmonary disease, a disorder of the peripheral or central nervous system or a cardiovascular disorder, comprising the step of:

administering to a patient in need thereof a therapeutically effective dose of a reagent that modulates a function of a human matriptase-like serine protease, whereby symptoms of the matriptase-like serine protease dysfunction related disease are ameliorated.

69. The method of claim 68 wherein the reagent is identified by the method of claim 36.

70. The method of claim 68 wherein the reagent is identified by the method of claim 45.

71. The method of claim 68 wherein the reagent is identified by the method of claim 49.

Fig. 1

aactccttttcctgcgggaacagccagtgtgtgaccaagg
tgaacccggagtgtgacgaccaggaggactgctccgatgg
gtccgacgaggcgcactgcgagtgtggcttgcagcctgcc
tggaggatggccggcaggatcgtgggcggcatggaagcat
ccccgggggagtttccgtggcaagccagccttcgagagaa
caaggagcacttctgtggggccgccatcatcaacgccagg
tggctggtgtctgctgctcactgcttcaatgagttccaag
acccgacgaagtgggtggcctacgtgggtgcgacctacct
cagcggctcggaggccagcaccgtgcgggcccaggtggtc
cagatcgtcaagcacccccctgtacaacgcggacacggccg
actttgacgtggctgtgctggagctgaccagcccctctgcc
tttcggccggcacatccagcccgtgtgcctcccggctgcc
acacacatctgtcccacccagcaagaagtgcctgatctcag
gctggggctacctcaaggaggacttccgtaagcatcttcc
tcggcctgcaatggtcaagccagaggtgctgcagaaagcc
actgtggagctgctggaccaggcactgtgtgccagcttgt
acggccattcactcactgacaggatggtgtgcgctggcta
cctggacgggaaggtggactcctgccagggtgactcagga
ggaccctggtctgcgaggagcccctctggcggttctttc
tggctggcatcgtgagctggggaatcgggtgtgcggaagc
ccggcgtccaggggtctatgcccgagtcaccaggctacgt
gactggatcctggaggccaccaccaaagccagcatgcctc
tggccccccatcatggctcctgccccctgccgagcacc
agcctggcccaccagtcctgagagccctgtggtcagcacc
cccaccaaatcgatgcaggcctcagtaccgtgcctcttg
actgggtcaccgttcctaagctacaag

Region covered by EST in bold

Fig. 2

NSFSCGNSQC VTKVNPECDD QEDCSDGSDE AHCECGLQPA
WRMAGRIVGG MEASPGEFPW QASLRENKEH FCGAAIINAR
WLVSAAHCFN EFQDPTKWVA YVGATYLSGS EASTVRAQVV
QIVKHPLYNA DTADFDVAVL ELTSPLPFGR HIQPVCLPAA
THIFPPSKKC LISGWGYLKE DFRKHLPRPA MVKPEVLQKA
TVELLDQALC ASLYGHSLTD RMVCAGYLDG KVDSCQGDSG
GPLVCEEPSG RFFLAGIVSW GIGCAEARRP GVYARVTRLR
DWILEATTKA SMPLAPTMAP APAAPSTAWP TSPESPVVST
PTKSMQALST VPLDWVTVPK LQ

Region covered by EST in bold

Fig. 3

MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGP
GRWVVLAAVLIGLLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDA
YENSNSTEFVSLASKVKDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYW
SEFSIPQHLVEEAERVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTVQR
TQDNSCSFGLHARGVELMRFTTPGFPDSPYPAHARCQWALRGDADSVLSLTF
RSFDLASCDERGSDLVTVYNTLSPMEPHALVQLCGTYPPSYNLTFHSSQNVL
LITLITNTERRHPGFEATFFQLPRMSSCGGRLRKAQGTFNSPYYPGHYPPNI
DCTWNIEVPNNQHVKVRFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQF
VVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTGRCIRKE
LRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCKPLFWVCDSVNDCGDNSDE
QGCSCPAQTFRCSNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVVTCTKHTY
RCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQARVVGGTDADEG
EWPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFL
GLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVR
PICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCENLL
PQQITPRMMCVGFLSGGVDSCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQ
RNKPGVYTRLPLFRDWIKENTGV

Fig. 4

IvGGreaqpgsfgsPwqvslqvrsgggsrkhfCGGsLisenwVLTAAHCvsga
asapass

vrVSlsvrlGehnlsltegteqkfdvkktiivHpnynpdtldngaYdnDiALl
kLkspgv

tlgdtvrpicLpsassdlpvGttctvsGwGrrptknlglsdtLqevvvpvv
sretCrsayeyggtdDkvefvtdnmiCagal.ggkdaCqGDSGGPLvcsdgnr
dgrwelv

GivSwGsygCargnkPGvytrVssyldWI

Fig. 5

stCggpdeFqCgsgrrCIprswvCDGdpDCeDGSDEslenCaa

Fig. 6

atggagcccactgtggctgacgtacacctcgtgcccaggacaaccaaggaagt
ccccgctctggatgccgcgtgctgtcgagcggccagcattggcgtggtggcca
ccagccttgtcgtcctcaccctgggagtcctttggccttcctctctacacag
ggcttccacgtggaccacacggccgagctgcggggaatccggtggaccagcag
tttgcggcgggagacctcggactatcaccgcacgctgacgcccaccctggagg
cactgctgcactttctgctgcgacccctccagacgctgagcctgggcctggag
gaggagctattgcagcgagggatccgggcaaggctgcgggagcacggcatctc
cctggctgcctatggcacaattgtgtcggctgagctcacagggagacataagg
gacccttggcagaaagagacttcaaatcaggccgctgtccagggaactccttt
tcctgcgggaacagccagtgtgtgaccaaggtgaacccggagtgtgacgacca
ggaggactgctccgatgggtccgacgaggcgcactgcgagtgtggcttgcagc
ctgcctggaggatggccggcaggatcgtgggcggcatggaagcatccccgggg
gagtttccgtggcaagccagccttcgagagaacaaggagcacttctgtggggc
cgccatcatcaacgccaggtggctggtgtctgctgctcactgcttcaatgagt
tccaagacccgacgaagtgggtggcctacgtgggtgcgacctacctcagcggc
tcggaggccagcaccgtgcgggcccaggtggtccagatcgtcaagcacccccct
gtacaacgcggacacggccgactttgacgtggctgtgctggagctgaccagcc
ctctgcctttcggccggcacatccagcccgtgtgcctcccggctgccacacac
atcttcccacccagcaagaagtgcctgatctcaggctggggctacctcaagga
ggacttcctggtcaagccagaggtgctgcagaaagccactgtggagctgctgg
accaggcactgtgtgccagcttgtacggccattcactcactgacaggatggtg
tgcgctggctacctggacgggaaggtggactcctgccagggtgactcaggagg
acccctggtctgcgaggagccctctggccggttctttctggctggcatcgtga
gctggggaatcgggtgtgcggaagcccggcgtccaggggtctatgcccgagtc
accaggctacgtgactggatcctggaggccaccaccaaagccagcatgcctct
ggcccccaccatggctcctgcccctgccgcccccagcacagcctggcccacca
gtcctgagagccctgtggtcagcacccccaccaaatcgatgcaggccctcagt
accgtgcctcttgactgggtcaccgttcctaagctacaagtaaaaaaagaaag
aaagtacgaggaactgacatatgcaggactgaaatgtgagccgtcacagccgg
tggccattaacaatagagccaagagccccagggacagcaactgtgggacacac
ctgaactgtctccaagcagaaatgaccagcatgcggatagccaggagtgggga
tctg

Fig. 7

```
BLASTP - Query = 151_ext_TR1 (SEQ ID NO:2);
Hit = trembl|AF118224|AF118224_1 (SEQ ID NO:3)

     This hit is scoring at : 2e-73 (expectation value)

     Alignment length (overlap) : 294

     Identities : 44 %

     Scoring matrix : BLOSUM62 (used to infer consensus pattern)

     Database searched : nrdb

 Q:       1 NSFSCGNSQCVTKVNPECDDQEDCSDGSDEAHCECGLQPAWRMAGRIVGGMEASPGEFPW
             ::::.C N. C::K NPECD.:EDCSDGSDE..C:CGL:.  R.A R:VGG.:A..GE:PW
 H:      570 HTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQA-RVVGGTDADEGEWPW

             QASLRE-NKEHFCGAAIINARWLVSAAHCFNE-----FQDPTKWVAYVGATYLSGSEAST
             Q.SL.. .: H.CGA::I:..WLVSAAHC: :     :.DPT:W.A::G.   S ..A.
             QVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQSQRSAPG

                                  TRYPSIN_HIS
```

Fig. 7 (continued)

```
VRAQ-VVQIVKHPLYNADTADFDVAVLELTSPLPFGRHIQPVCLPAATHIFPPSKKCLIS
V:.: : :I:.HP.:N  T D:D:A:LEL..P..:.. ::P:CLP A:H:FP..K.. ::
VQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIWVT

                    ACT_SITE _ASP

GWGYLKEDFRKHLPRPAMVKPEVLQKATVELLDQALCASLYGHSLTDRMVCAGYLDGKVD
GWG:.: . .  L          :LQK.:.:.::Q..C.:L. ..:T.RM:C.G:L.G VD
GWGHTQYGGTGAL ----ILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGVD

SCQGDSGGPLVCEEPSGRFFLAGIVSWGIGCAEARRPGVYARVTRLRDWILEAT      287
SCQGDSGGPL . E..GR.F AG:VSWG GCA:..:PGVY.R:. .RDWI E T
SCQGDSGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENT      853

TRYPSIN_SER
```

BLOCKS underlined

Fig. 8

Prosite search results

Access#	From->To	Name	Doc#
PS00134	83->89	TRYPSIN_HIS	PDOC00124
PS00135	233->245	TRYPSIN_SER	PDOC00124

Fig. 9

BLOCKS search results

AC#	Description	Strength	Score
BL01253G	Type I fibronectin domain proteins.	1641	1537
AA#	231 vDsCQGDSGGPLVC (SEQ ID NO:4)		
BL00134A	Serine proteases, trypsin family, histidine p	1500	1415
AA#	71 CGAAIINARWLVSAAHC (SEQ ID NO:5)		
BL01253H	Type I fibronectin domain proteins.	1765	1409
AA#	251 ffLaGIvSWGiGCaEarrPGVYaRVTrlrDWIlEa (SEQ ID NO:6)		
BL00021D	Kringle domain proteins.	1556	1321
AA#	241 plvceEpsGRffLaGIVSWGiGCAEaRRPGVYARVTRLrDWI (SEQ ID NO:7)		
BL00021B	Kringle domain proteins.	1547	1318
AA#	71 CGAaiInaRWlVSAAHCF (SEQ ID NO:8)		
BL00495O	Apple domain proteins.	1756	1314
AA#	260 GiGCAearRPGVYarVtrlrDWILEaTtk (SEQ ID NO:9)		
BL01209	LDL-receptor class A (LDLRA) domain proteins	1413	1289

Fig. 9 (continued)

```
    AA#     17 CDdQEDCSDGSDE    (SEQ ID NO:10)

BL00134C      Serine proteases, trypsin family,          1245   1278
                   histidine p

    AA#    269 PGVYARVTRLRDWI    (SEQ ID NO:11)

BL00134B      Serine proteases, trypsin family,          1289   1264
                   histidine p

    AA#    232 DSCQGDSGGPLVCEEPSGRFFLAG (SEQ ID NO:12)

BL00495N      Apple domain proteins.                     1945   1249

    AA#    224 AGYldGkvDsCqGDSGGPLVCeepsgrfflagivs (SEQ ID NO:13)

BL01253E      Type I fibronectin domain proteins.        1661   1220

    AA#    143 SplpfgrhiQPVCLPaAthiFPpskKClISGWGyLKE (SEQ ID NO:14)
```

Fig. 10

```
HMMPFAM - alignment of 151_ext_TR1 against pfam|hmm|trypsin (SEQ ID NO:15)
     Trypsin
     This hit is scoring at : 287.5; Expect = 1.8e-90
     Scoring matrix : BLOSUM62 (used to infer consensus pattern)
 Q:      47 IVGGMEASPGEF--PWQASLRENK- EHFCGAAIINARWLVSAAHCFNE--FQDPTK
            IVGG.EA.PG.F  PWQ.SL: ..     :HFCG.::I:..W:::AAHC.:    ...:.
 H:       1 IvGGreaqpgsfgsPwqvslqvrsgggsrkhfCGGsLisenwVLTAAHCvsgaasapass
            WVA----YVGATYLSGSEASTVRAQVVQ-IVKHPLYNADTADFDVAVLELTSP-L
              .     :G.  LS :E.:. : .V : I: HP YN.DT.D      D:A:L:L.SP :
            vrVSlsvrlGehnlsltegteqkfdvkktiivHpnynpdtldngaYdnDiALlkLkspgv

            PFGRHIQPVCLPAATHIFPPSKKCLISGWGY-LkEDFRKhlprpamvKPEVLQKATVELL
            ..G  ::P:CLP:A:. .P ...C.:SGWG    ..             .:.LQ:..V.::
            tlgdtvrpicLpsassdlpvGttctvsGwGrrp.tknlg.......lsdtLqevvvpvv

            DQALCASL-YGHS    -LTDRMVCAGYLdGKVDSCQGDSGGPLVCEEPS-GRFFLA
            .:..C.S.  YG :      :TD.M:CAG L G  D:CQGDSGGPLVC.: :  GR: L.
            sretCrsayeyggtdDkvefvtdnmiCagal.ggkdaCqGDSGGPLvcsdgnrdgrwelv

            GIVSWG IGCAEARRPGVYARVTRLRDWI     283
            GIVSWG .GCA...:PGVY.RV:.. DWI
            GivSwGsygCargnkPGvytrVssyldWI     259
```

Fig. 11

```
HMMPFAM - alignment of 151_ext_TR1 against pfam|hmm|ldl_recept_a

    Low-density lipoprotein receptor domain (SEQ ID NO:16)

    This hit is scoring at : 10.9; Expect = 0.061

    Scoring matrix : BLOSUM62 (used to infer consensus pattern)

Q:       1  NSFSCGNSQ-CVTKvNPECDDQEDCSDGSDE--AHCEC      35
                 :.F.CG:.: C:.: :  CD...DC.DGSDE  .:C..
H:       1 stCggpdeFqCgsgrrCIpr.swvCDGdpDCeDGSDEslenCaa      43
```

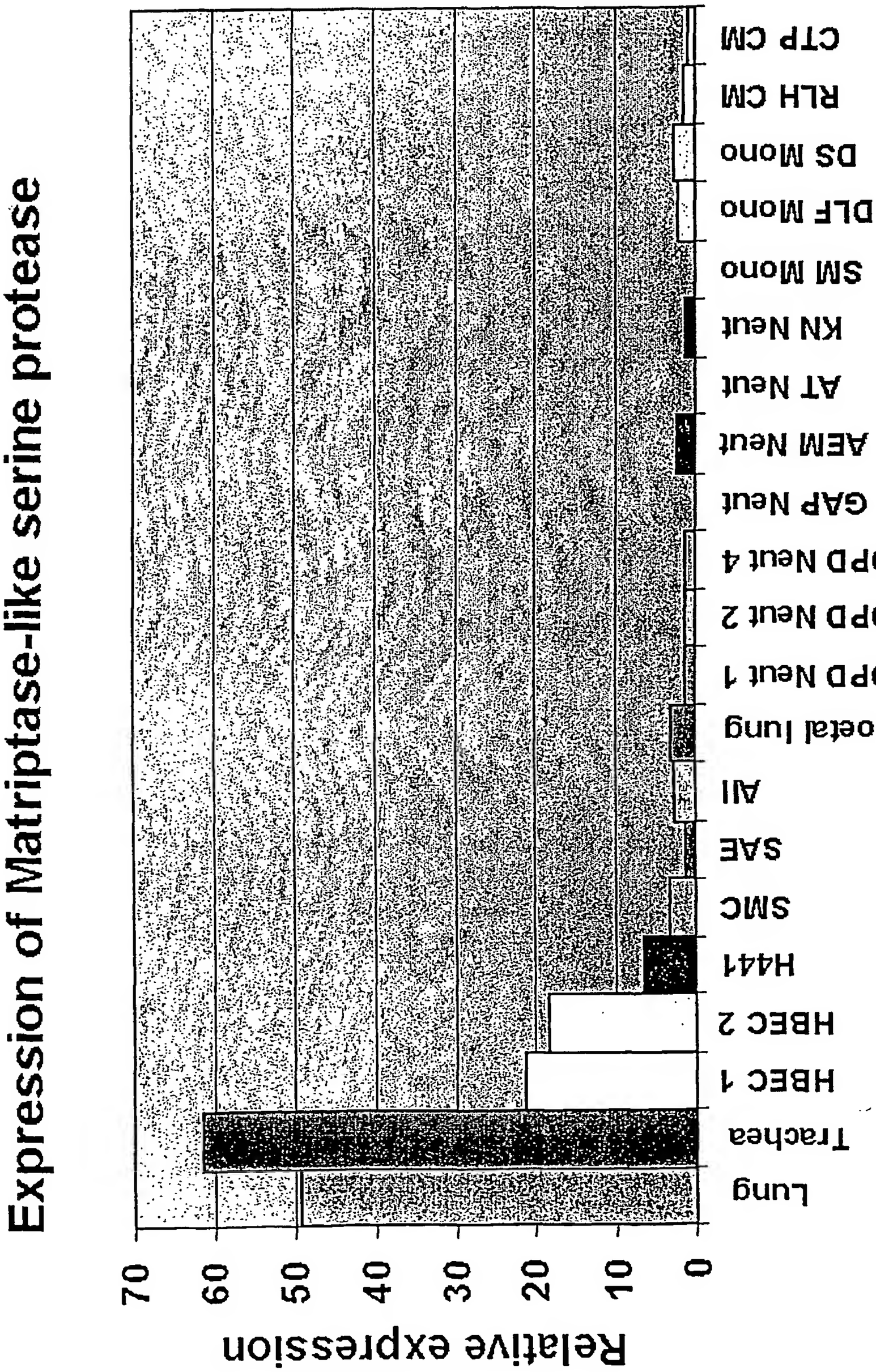
Expression of Matriptase-like serine protease
Relative expression
0
10
20
30
40
50
60
70
Lung
Trachea
HBEC 1
HBEC 2
PBLC 2
H441
SMC
SAE
AII
Foetal lung
COPD Neut 1
COPD Neut 2
COPD Neut 4
GAP Neut
AEM Neut
AT Neut
KN Neut
SM Mono
DLF Mono
DS Mono
RLH CM
CTP CM

Fig. 12

Fig. 13

Expression of Matriptase-like serine protease II

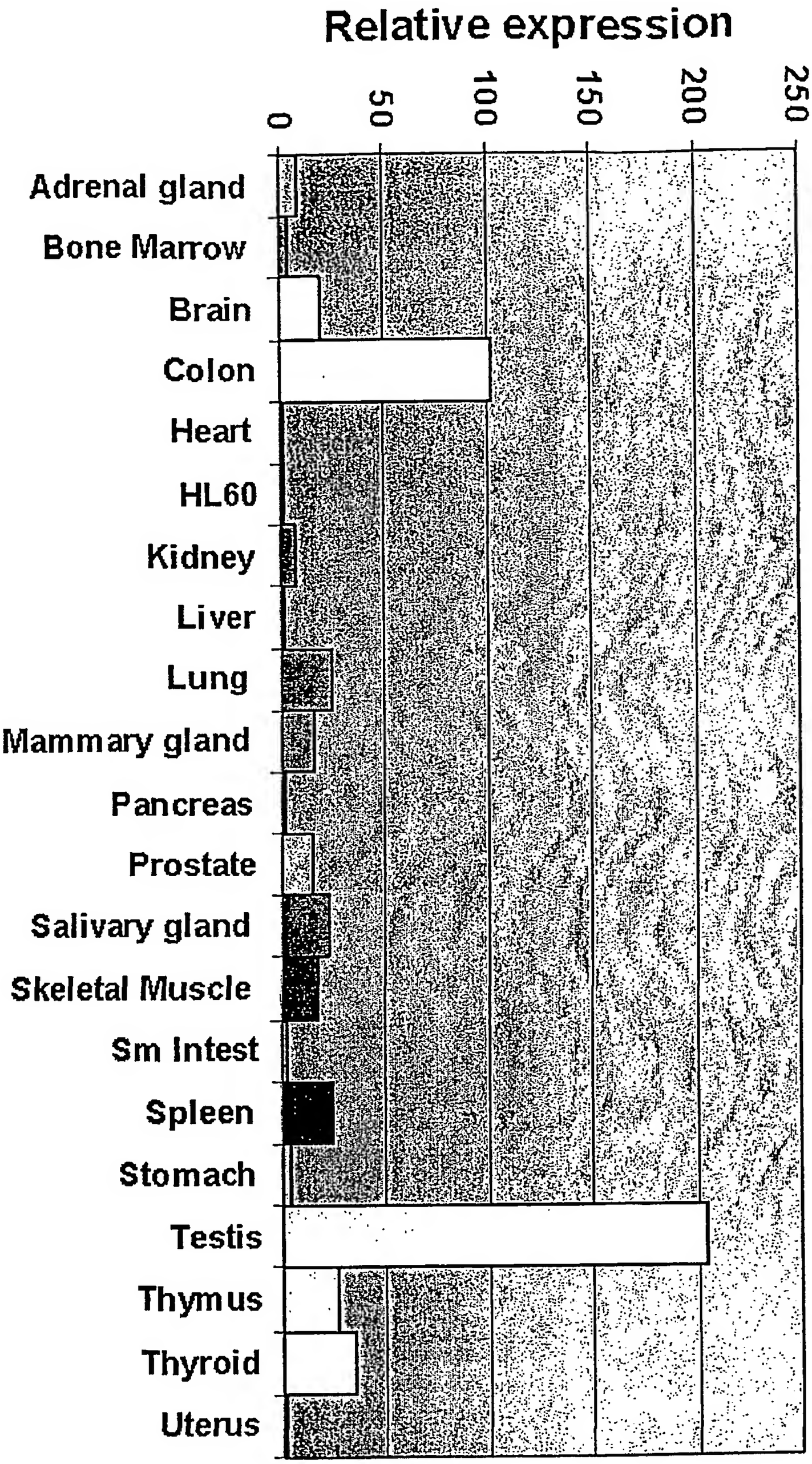

Fig. 14

```
BLASTP - alignment of 151_ext2_protein against trembl|AF118224|AF118224_1
    This hit is scoring at : 3e-74 (expectation value)
    Alignment length (overlap) : 289
    Identities : 45 %
    Scoring matrix : BLOSUM62 (used to infer consensus pattern)
    Database searched : nrdb_1_;
Q:    154 CPGNSFSCGNSQCVTKVNPECDDQEDCSDGSDEAHCECGLQPAWRMAGRIVGGMEASPGE
          C. ::::.C N. C::K NPECD.:EDCSDGSDE..C:CGL:.  R.A R:VGG.:A..GE
H:    567 CTKHTYRCLNGLCLSKGNPECDGKEDCSDGSDEKDCDCGLRSFTRQA-RVVGGTDADEGE

FPWQASLRE-NKEHFCGAAIINARWLVSAAHCFNE-----FQDPTKWVAYVGATYLSGSE
:PWQ.SL.. .: H.CGA::I:..WLVSAAHC: :      :.DPT:W.A::G.   S ..
WPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSDPTQWTAFLGLHDQSQRS
                          TRYPSIN_HIS

ASTVRAQ-VVQIVKHPLYNADTADFDVAVLELTSPLPFGRHIQPVCLPAATHIFPPSKKC
A. V:.: : :I:.HP.:N  T D:D:A:LEL..P..:.. ::P:CLP A:H:FP..K..
APGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAI
                    ACT_SITE _ASP

LISGWGYLKEDFLVKPEVLQKATVELLDQALCASLYGHSLTDRMVCAGYLDGKVDSCQGD
 ::GWG:.: .  . . :LQK..:.:::Q..C.:L. ..:T.RM:C.G:L.G VDSCQGD
```

Fig 14 (continued)

```
        WVTGWGHTQYGG-TGALILQKGEIRVINQTTCENLLPQQITPRMMCVGFLSGGVDSCQGD

        SGGPLVCEEPSGRFFLAGIVSWGIGCAEARRPGVYARVTRLRDWILEAT      435
        SGGPL . E..GR.F AG:VSWG GCA:..:PGVY.R:. .RDWI E T
        SGGPLSSVEADGRIFQAGVVSWGDGCAQRNKPGVYTRLPLFRDWIKENT      853
    TRYPSIN_SER

This hit is scoring at : 0.077 (expectation value)
Alignment length (overlap) : 41
Identities : 43 %
    Scoring matrix : BLOSUM62 (used to infer consensus pattern)
    Database searched : nrdb_1_;
Q:     151 SGRCPGNSFSCGNSQCVTKVNPECDDQEDCSDGSDEAHCEC      191
           S. CPG. F:C ..:C:.K . .CD. .DC:D SDE.:C.C
H:     450 SDPCPGQ-FTCRTGRCIRK-ELRCDGWADCTDHSDELNCSC      488
```

Fig. 15

BLOCKS search results are same as the previous version of 151:

AC#		Description	Strength	Score
BL01253G	0	Type I fibronectin domain proteins.	1453	1660
AA#	379	vDsCQGDSGGPLVC		
BL01253H	0	Type I fibronectin domain proteins.	1520	1456
AA#	399	ffLaGIvSWGiGCaEarrPGVYaRVTrlrDWIlEa		
BL00134A	0	Serine proteases, trypsin family, histidine p	1439	1432
AA#	227	CGAAIINARWLVSAAHC		
BL00495O	0	Apple domain proteins.	1396	1388
AA#	408	GiGCAearRPGVYarVtrlrDWILEaTtk		
BL00021B	0	Kringle domain proteins.	1466	1374
AA#	227	CGAaiInaRWlVSAAHCF		
BL00021D	0	Kringle domain proteins.	1410	1350
AA#	389	plvceEpsGRffLaGIVSWGiGCAEaRRPGVYARVTRLrDWI		
BL01209	0	LDL-receptor class A (LDLRA) domain proteins.	1296	1316
AA#	173	CDdQEDCSDGSDE		
BL00495N	0	Apple domain proteins.	1538	1300
AA#	372	AGYldGkvDsCqGDSGGPLVCeepsgrfflagivs		
BL00134C	0	Serine proteases, trypsin family, histidine p	1163	1298
AA#	417	PGVYARVTRLRDWI		
BL00134B	0	Serine proteases, trypsin family, histidine p	1225	1288
AA#	380	DSCQGDSGGPLVCEEPSGRFFLAG		
BL01253E	0	Type I fibronectin domain proteins.	1385	1261
AA#	299	SplpfgrhiQPVCLPaAthiFPpskKClISGWGyLKE		

Fig. 16

```
HMMPFAM - alignment of 151_ext2_protein against pfam|hmm|trypsin
    Trypsin
    This hit is scoring at : 295.9; Expect = 3.8e-93
    Scoring matrix : BLOSUM62 (used to infer consensus pattern)
Q:    203 IVGGMEASPGEF--PWQASLRENK-----EHFCGAAIINARWLVSAAHCFNE--FQDPTK
          IVGG.EA.PG.F  PWQ.SL: ..      :HFCG.::I:..W:::AAHC.:    ...:.
H:      1 IvGGreaqpgsfgsPwqvslqvrsgggsrkhfCGGsLisenwVLTAAHCvsgaasapass

          WVA----YVGATYLSGSEASTVRAQVVQ-IVKHPLYNADTAD-----FDVAVLELTSP-L
            .      :G.  LS :E.:. : .V : I: HP YN.DT.D      D:A:L:L.SP :
          vrVSlsvrlGehnlsltegteqkfdvkktiivHpnynpdtldngaYdnDiALlkLkspgv

          PFGRHIQPVCLPAATHIFPPSKKCLISGWGY--LKEDFlvKPEVLQKATVELLDQALCAS
          ..G  ::P:CLP:A:. .P ...C.:SGWG   .K.      .:.LQ:..V.:::.:.C.S
          tlgdtvrpicLpsassdlpvGttctvsGwGrrptknlg..lsdtLqevvvpvvsretCrs

          L--YGHS------LTDRMVCAGYLdGKVDSCQGDSGGPLVCEEPS--GRFFLAGIVSWG-
          .  YG :       :TD.M:CAG L G  D:CQGDSGGPLVC.: :  GR: L.GIVSWG
          ayeyggtdDkvefvtdnmiCagal.ggkdaCqGDSGGPLvcsdgnrdgrwelvGivSwGs

          IGCAEARRPGVYARVTRLRDWI     431
          .GCA...:PGVY.RV:.. DWI
          ygCargnkPGvytrVssyldWI     259
```

Fig. 17

```
HMMPFAM - alignment of 151_ext2_protein against pfam|hmm|ldl_recept_a
Low-density lipoprotein receptor doma

This hit is scoring at : 26.5; Expect = 0.00064
Scoring matrix : BLOSUM62 (used to infer consensus pattern)

Q:     152 GRCP-GNSFSCGNSQ-CVTKvNPECDDQEDCSDGSDE--AHCEC     191
           ..C   :.F.CG:.: C:.: :  CD...DC.DGSDE  .:C..
H:       1 stCggpdeFqCgsgrrCIpr.swvCDGdpDCeDGSDEslenCaa      43

Here is the exon/intron sturcture in alignment with genomic sequence:
Q = LBRI_151_ext2_DNA; H = AC011542.5

exon 1
Q:       1 atggagcccactgtggctgacgtacacctcgtgcccaggacaaccaaggaagtccccgct
H:   10873 atggagcccactgtggctgacgtacacctcgtgcccaggacaaccaaggaagtccccgct

           ctggatgccgcgtgctgtcgagcggccagcattggcgtggtggccaccagccttgtcgtc
           ctggatgccgcgtgctgtcgagcggccagcattggcgtggtggccaccagccttgtcgtc

           ctcaccctgggagtcctttgg      142
           ctcaccctgggagtcctttgg(gta    11014

exon 2
Q:     143      ccttcctctctacacagggcttccacgtggaccacacggccgagctgcggggaatccgg
H:   17744 cag)ccttcctctctacacagggcttccacgtggaccacacggccgagctgcggggaatccgg

           tggaccagcagtttgcggcgggagacctcggactatcaccgcacgctgacgcccaccctg
           tggaccagcagtttgcggcgggagacctcggactatcaccgcacgctgacgcccaccctg
```

Fig. 17 (continued)

```
          gaggcactg        270
          gaggcactg(gtg    17871


exon 3
Q:     271     ctgcactttctgctgcgacccctccagacgctgagcctgggcctggaggaggagctatt
H:   20257 cag)ctgcactttctgctgcgacccctccagacgctgagcctgggcctggaggaggagctatt

          gcagcgagggatccgggcaaggctgcgggagcacggcatctccctggctgcctatggcac
          gcagcgagggatccgggcaaggctgcgggagcacggcatctccctggctgcctatggcac

          aattgtgtcggctgagctcacag     412
          aattgtgtcggctgagctcacag(gtg   20398

exon 4
Q:     413     ggagacataagggacccttggcagaaagagacttcaaatcag       454
H:   23180 cag)ggagacataagggacccttggcagaaagagacttcaaatcag(gta    23221

exon 5
Q:     455      gccgctgtccagggaactccttttcctgcgggaacagccagtgtgtgaccaaggtg
H:   24287 tcag)gccgctgtccagggaactccttttcctgcgggaacagccagtgtgtgaccaaggtg

           aacccggagtgtgacgaccaggaggactgctccgatgggtccgacgaggcgcactgcg      568
           aacccggagtgtgacgaccaggaggactgctccgatgggtccgacgaggcgcactgcg(gtc    24400

exon 6
Q:     569     agtgtggcttgcagcctgcctggaggatggccggcaggatcgtgggcggcatggaagca
H:   26579 cag)agtgtggcttgcagcctgcctggaggatggccggcaggatcgtgggcggcatggaagca
```

Fig. 17 (continued)

```
          tccccgggggagtttccgtggcaagccagccttcgagagaacaaggagcacttctgtggg
          tccccgggggagtttccgtggcaagccagccttcgagagaacaaggagcacttctgtggg

          gccgccatcatcaacgccaggtggctggtgtctgctgctcactgcttcaatga      740
          gccgccatcatcaacgccaggtggctggtgtctgctgctcactgcttcaatga(gta   26750



exon 7
Q:     741     gttccaagacccgacgaagtgggtggcctacgtgggtgcgacctacctcagcggctcgga
H:   29675 cag)gttccaagacccgacgaagtgggtggcctacgtgggtgcgacctacctcagcggctcgga
          ggccagcaccgtgcgggcccaggtggtccagatcgtcaagcacccccctgtacaacgcgga
          ggccagcaccgtgcgggcccaggtggtccagatcgtcaagcacccccctgtacaacgcgga

          cacggccgactttgacgtggctgtgctggagctgaccagccctctgcctttcggccggca
          cacggccgactttgacgtggctgtgctggagctgaccagccctctgcctttcggccggca

          catccagcccgtgtgcctcccggctgccacacacatcttcccacccagcaagaagtgcct
          catccagcccgtgtgcctcccggctgccacacacatcttcccacccagcaagaagtgcct

          gatctcaggctggggctacctcaaggaggacttcc    1015
          gatctcaggctggggctacctcaaggaggacttcc(gta   29949

exon 8
Q:    1016     tggtcaagccagaggtgctgcagaaagccactgtggagctgctggaccaggcactgtgtg
H:   31577 cag)tggtcaagccagaggtgctgcagaaagccactgtggagctgctggaccaggcactgtgtg
```

Fig. 17 (continued)

```
          ccagcttgtacggccattcactcactgacaggatggtgtgcgctggctacctggacggga
          ccagcttgtacggccattcactcactgacaggatggtgtgcgctggctacctggacggga

          aggtggactcctgccag    1152
          aggtggactcctgccag(gtga   31713

exon 9
Q:   1153     ggtgactcaggaggacccctggtctgcgaggagccctctggccggttctttctggctg
H:  35018 tag)ggtgactcaggaggacccctggtctgcgaggagccctctggccggttctttctggctg

          gcatcgtgagctggggaatcgggtgtgcggaagcccggcgtccaggggtctatgcccgag
          gcatcgtgagctggggaatcgggtgtgcggaagcccggcgtccaggggtctatgcccgag
```

Fig. 17 (continued)

```
tcaccaggctacgtgactggatcctggaggccaccaccaaagccagcatgcctctggccc
tcaccaggctacgtgactggatcctggaggccaccaccaaagccagcatgcctctggccc

ccaccatggctcctgcccctgccgccccagcacagcctggcccaccagtcctgagagcc
ccaccatggctcctgcccctgccgccccagcacagcctggcccaccagtcctgagagcc

ctgtggtcagcacccccaccaaatcgatgcaggccctcagtaccgtgcctcttgactggg
ctgtggtcagcacccccaccaaatcgatgcaggccctcagtaccgtgcctcttgactggg

tcaccgttcctaagctacaag    1471
tcaccgttcctaagctacaag(gta   35336

exon 10
Q:   1472     taaaaaaagaaagaaagtacgaggaactgacatatgcaggactgaaatgtgagccgtca
H:  36697 cag)taaaaaaagaaagaaagtacgaggaactgacatatgcaggactgaaatgtgagccgtca

          cagccggtggccattaacaatagagccaagagccccagggacagcaactgtgggacacac
          cagccggtggccattaacaatagagccaagagccccagggacagcaactgtgggacacac

          ctgaactgtctccaagcagaaatgaccagcatgcggatagccaggagtggggatctg    1646
          ctgaactgtctccaagcagaaatgaccagcatgcggatagccaggagtggggatctg   36871
```

SEQUENCE LISTING

<110> Bayer AG

<120> REGULATION OF HUMAN MATRIPTASE-LIKE SERINE PROTEASE

<130> LIO127 Foreign Countries

<150> US 60/220,807
<151> 2000-07-25

<150> US 60/280,109
<151> 2001-04-02

<160> 18

<170> PatentIn version 3.1

<210> 1
<211> 1027
<212> DNA
<213> Homo sapiens

<400> 1
aactcctttt cctgcgggaa cagccagtgt gtgaccaagg tgaacccgga gtgtgacgac 60

caggaggact gctccgatgg gtccgacgag gcgcactgcg agtgtggctt gcagcctgcc 120

tggaggatgg ccggcaggat cgtgggcggc atggaagcat ccccggggga gtttccgtgg 180

caagccagcc ttcgagagaa caaggagcac ttctgtgggg ccgccatcat caacgccagg 240

tggctggtgt ctgctgctca ctgcttcaat gagttccaag acccgacgaa gtgggtggcc 300

tacgtgggtg cgacctacct cagcggctcg gaggccagca ccgtgcgggc ccaggtggtc 360

cagatcgtca agcacccccct gtacaacgcg gacacggccg actttgacgt ggctgtgctg 420

gagctgacca gccctctgcc tttcggccgg cacatccagc ccgtgtgcct cccggctgcc 480

acacacatct tcccacccag caagaagtgc ctgatctcag gctggggcta cctcaaggag 540

gacttccgta agcatcttcc tcggcctgca atggtcaagc cagaggtgct gcagaaagcc 600

actgtggagc tgctggacca ggcactgtgt gccagcttgt acggccattc actcactgac 660

aggatggtgt gcgctggcta cctggacggg aaggtggact cctgccaggg tgactcagga 720

ggacccctgg tctgcgagga gccctctggc cggttctttc tggctggcat cgtgagctgg 780

ggaatcgggt gtgcggaagc ccggcgtcca ggggtctatg cccgagtcac caggctacgt 840

gactggatcc tggaggccac caccaaagcc agcatgcctc tggcccccac catggctcct 900

gcccctgccg cccccagcac agcctggccc accagtcctg agagccctgt ggtcagcacc 960

cccaccaaat cgatgcaggc cctcagtacc gtgcctcttg actgggtcac cgttcctaag 1020

ctacaag 1027

<210> 2
<211> 342

<212> PRT
<213> Homo sapiens

<400> 2

Asn Ser Phe Ser Cys Gly Asn Ser Gln Cys Val Thr Lys Val Asn Pro
1 5 10 15

Glu Cys Asp Asp Gln Glu Asp Cys Ser Asp Gly Ser Asp Glu Ala His
20 25 30

Cys Glu Cys Gly Leu Gln Pro Ala Trp Arg Met Ala Gly Arg Ile Val
35 40 45

Gly Gly Met Glu Ala Ser Pro Gly Glu Phe Pro Trp Gln Ala Ser Leu
50 55 60

Arg Glu Asn Lys Glu His Phe Cys Gly Ala Ala Ile Ile Asn Ala Arg
65 70 75 80

Trp Leu Val Ser Ala Ala His Cys Phe Asn Glu Phe Gln Asp Pro Thr
85 90 95

Lys Trp Val Ala Tyr Val Gly Ala Thr Tyr Leu Ser Gly Ser Glu Ala
100 105 110

Ser Thr Val Arg Ala Gln Val Val Gln Ile Val Lys His Pro Leu Tyr
115 120 125

Asn Ala Asp Thr Ala Asp Phe Asp Val Ala Val Leu Glu Leu Thr Ser
130 135 140

Pro Leu Pro Phe Gly Arg His Ile Gln Pro Val Cys Leu Pro Ala Ala
145 150 155 160

Thr His Ile Phe Pro Pro Ser Lys Lys Cys Leu Ile Ser Gly Trp Gly
165 170 175

Tyr Leu Lys Glu Asp Phe Arg Lys His Leu Pro Arg Pro Ala Met Val
180 185 190

Lys Pro Glu Val Leu Gln Lys Ala Thr Val Glu Leu Leu Asp Gln Ala
195 200 205

Leu Cys Ala Ser Leu Tyr Gly His Ser Leu Thr Asp Arg Met Val Cys
210 215 220

Ala Gly Tyr Leu Asp Gly Lys Val Asp Ser Cys Gln Gly Asp Ser Gly
225 230 235 240

Gly Pro Leu Val Cys Glu Glu Pro Ser Gly Arg Phe Phe Leu Ala Gly
245 250 255

Ile Val Ser Trp Gly Ile Gly Cys Ala Glu Ala Arg Arg Pro Gly Val
260 265 270

Tyr Ala Arg Val Thr Arg Leu Arg Asp Trp Ile Leu Glu Ala Thr Thr
275 280 285

Lys Ala Ser Met Pro Leu Ala Pro Thr Met Ala Pro Ala Pro Ala Ala
290 295 300

Pro Ser Thr Ala Trp Pro Thr Ser Pro Glu Ser Pro Val Val Ser Thr
305 310 315 320

Pro Thr Lys Ser Met Gln Ala Leu Ser Thr Val Pro Leu Asp Trp Val
325 330 335

Thr Val Pro Lys Leu Gln
340

<210> 3
<211> 855
<212> PRT
<213> Homo sapiens

<400> 3

Met Gly Ser Asp Arg Ala Arg Lys Gly Gly Gly Gly Pro Lys Asp Phe
1 5 10 15

Gly Ala Gly Leu Lys Tyr Asn Ser Arg His Glu Lys Val Asn Gly Leu
20 25 30

Glu Glu Gly Val Glu Phe Leu Pro Val Asn Asn Val Lys Lys Val Glu
35 40 45

Lys His Gly Pro Gly Arg Trp Val Val Leu Ala Ala Val Leu Ile Gly
50 55 60

Leu Leu Leu Val Leu Leu Gly Ile Gly Phe Leu Val Trp His Leu Gln
65 70 75 80

Tyr Arg Asp Val Arg Val Gln Lys Val Phe Asn Gly Tyr Met Arg Ile
85 90 95

Thr Asn Glu Asn Phe Val Asp Ala Tyr Glu Asn Ser Asn Ser Thr Glu
100 105 110

```
Phe Val Ser Leu Ala Ser Lys Val Lys Asp Ala Leu Lys Leu Leu Tyr
        115                 120                 125

Ser Gly Val Pro Phe Leu Gly Pro Tyr His Lys Glu Ser Ala Val Thr
    130                 135                 140

Ala Phe Ser Glu Gly Ser Val Ile Ala Tyr Tyr Trp Ser Glu Phe Ser
145                 150                 155                 160

Ile Pro Gln His Leu Val Glu Glu Ala Glu Arg Val Met Ala Glu Glu
                165                 170                 175

Arg Val Val Met Leu Pro Pro Arg Ala Arg Ser Leu Lys Ser Phe Val
            180                 185                 190

Val Thr Ser Val Val Ala Phe Pro Thr Asp Ser Lys Thr Val Gln Arg
        195                 200                 205

Thr Gln Asp Asn Ser Cys Ser Phe Gly Leu His Ala Arg Gly Val Glu
    210                 215                 220

Leu Met Arg Phe Thr Thr Pro Gly Phe Pro Asp Ser Pro Tyr Pro Ala
225                 230                 235                 240

His Ala Arg Cys Gln Trp Ala Leu Arg Gly Asp Ala Asp Ser Val Leu
                245                 250                 255

Ser Leu Thr Phe Arg Ser Phe Asp Leu Ala Ser Cys Asp Glu Arg Gly
            260                 265                 270

Ser Asp Leu Val Thr Val Tyr Asn Thr Leu Ser Pro Met Glu Pro His
        275                 280                 285

Ala Leu Val Gln Leu Cys Gly Thr Tyr Pro Pro Ser Tyr Asn Leu Thr
    290                 295                 300

Phe His Ser Ser Gln Asn Val Leu Leu Ile Thr Leu Ile Thr Asn Thr
305                 310                 315                 320

Glu Arg Arg His Pro Gly Phe Glu Ala Thr Phe Phe Gln Leu Pro Arg
                325                 330                 335

Met Ser Ser Cys Gly Gly Arg Leu Arg Lys Ala Gln Gly Thr Phe Asn
            340                 345                 350

Ser Pro Tyr Tyr Pro Gly His Tyr Pro Pro Asn Ile Asp Cys Thr Trp
```

```
     355                 360                 365

Asn Ile Glu Val Pro Asn Asn Gln His Val Lys Val Arg Phe Lys Phe
    370                 375                 380

Phe Tyr Leu Leu Glu Pro Gly Val Pro Ala Gly Thr Cys Pro Lys Asp
385                 390                 395                 400

Tyr Val Glu Ile Asn Gly Glu Lys Tyr Cys Gly Glu Arg Ser Gln Phe
                405                 410                 415

Val Val Thr Ser Asn Ser Asn Lys Ile Thr Val Arg Phe His Ser Asp
            420                 425                 430

Gln Ser Tyr Thr Asp Thr Gly Phe Leu Ala Glu Tyr Leu Ser Tyr Asp
        435                 440                 445

Ser Ser Asp Pro Cys Pro Gly Gln Phe Thr Cys Arg Thr Gly Arg Cys
    450                 455                 460

Ile Arg Lys Glu Leu Arg Cys Asp Gly Trp Ala Asp Cys Thr Asp His
465                 470                 475                 480

Ser Asp Glu Leu Asn Cys Ser Cys Asp Ala Gly His Gln Phe Thr Cys
                485                 490                 495

Lys Asn Lys Phe Cys Lys Pro Leu Phe Trp Val Cys Asp Ser Val Asn
            500                 505                 510

Asp Cys Gly Asp Asn Ser Asp Glu Gln Gly Cys Ser Cys Pro Ala Gln
        515                 520                 525

Thr Phe Arg Cys Ser Asn Gly Lys Cys Leu Ser Lys Ser Gln Gln Cys
    530                 535                 540

Asn Gly Lys Asp Asp Cys Gly Asp Gly Ser Asp Glu Ala Ser Cys Pro
545                 550                 555                 560

Lys Val Asn Val Val Thr Cys Thr Lys His Thr Tyr Arg Cys Leu Asn
                565                 570                 575

Gly Leu Cys Leu Ser Lys Gly Asn Pro Glu Cys Asp Gly Lys Glu Asp
            580                 585                 590

Cys Ser Asp Gly Ser Asp Glu Lys Asp Cys Asp Cys Gly Leu Arg Ser
        595                 600                 605
```

```
Phe Thr Arg Gln Ala Arg Val Val Gly Gly Thr Asp Ala Asp Glu Gly
    610                 615                 620

Glu Trp Pro Trp Gln Val Ser Leu His Ala Leu Gly Gln Gly His Ile
625                 630                 635                 640

Cys Gly Ala Ser Leu Ile Ser Pro Asn Trp Leu Val Ser Ala Ala His
                645                 650                 655

Cys Tyr Ile Asp Asp Arg Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp
            660                 665                 670

Thr Ala Phe Leu Gly Leu His Asp Gln Ser Gln Arg Ser Ala Pro Gly
        675                 680                 685

Val Gln Glu Arg Arg Leu Lys Arg Ile Ile Ser His Pro Phe Phe Asn
    690                 695                 700

Asp Phe Thr Phe Asp Tyr Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro
705                 710                 715                 720

Ala Glu Tyr Ser Ser Met Val Arg Pro Ile Cys Leu Pro Asp Ala Ser
                725                 730                 735

His Val Phe Pro Ala Gly Lys Ala Ile Trp Val Thr Gly Trp Gly His
            740                 745                 750

Thr Gln Tyr Gly Gly Thr Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile
        755                 760                 765

Arg Val Ile Asn Gln Thr Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile
    770                 775                 780

Thr Pro Arg Met Met Cys Val Gly Phe Leu Ser Gly Gly Val Asp Ser
785                 790                 795                 800

Cys Gln Gly Asp Ser Gly Gly Pro Leu Ser Ser Val Glu Ala Asp Gly
                805                 810                 815

Arg Ile Phe Gln Ala Gly Val Val Ser Trp Gly Asp Gly Cys Ala Gln
            820                 825                 830

Arg Asn Lys Pro Gly Val Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp
        835                 840                 845

Ile Lys Glu Asn Thr Gly Val
    850                 855
```

<210> 4
<211> 14
<212> PRT
<213> Homo sapiens

<400> 4

Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys
1 5 10

<210> 5
<211> 17
<212> PRT
<213> Homo sapiens

<400> 5

Cys Gly Ala Ala Ile Ile Asn Ala Arg Trp Leu Val Ser Ala Ala His
1 5 10 15

Cys

<210> 6
<211> 35
<212> PRT
<213> Homo sapiens

<400> 6

Phe Phe Leu Ala Gly Ile Val Ser Trp Gly Ile Gly Cys Ala Glu Ala
1 5 10 15

Arg Arg Pro Gly Val Tyr Ala Arg Val Thr Arg Leu Arg Asp Trp Ile
20 25 30

Leu Glu Ala
35

<210> 7
<211> 42
<212> PRT
<213> Homo sapiens

<400> 7

Pro Leu Val Cys Glu Glu Pro Ser Gly Arg Phe Phe Leu Ala Gly Ile
1 5 10 15

Val Ser Trp Gly Ile Gly Cys Ala Glu Ala Arg Arg Pro Gly Val Tyr
20 25 30

Ala Arg Val Thr Arg Leu Arg Asp Trp Ile
35 40

<210> 8
<211> 18
<212> PRT
<213> Homo sapiens

<400> 8

Cys Gly Ala Ala Ile Ile Asn Ala Arg Trp Leu Val Ser Ala Ala His
1 5 10 15

Cys Phe

<210> 9
<211> 29
<212> PRT
<213> Homo sapiens

<400> 9

Gly Ile Gly Cys Ala Glu Ala Arg Arg Pro Gly Val Tyr Ala Arg Val
1 5 10 15

Thr Arg Leu Arg Asp Trp Ile Leu Glu Ala Thr Thr Lys
20 25

<210> 10
<211> 13
<212> PRT
<213> Homo sapiens

<400> 10

Cys Asp Asp Gln Glu Asp Cys Ser Asp Gly Ser Asp Glu
1 5 10

<210> 11
<211> 14
<212> PRT
<213> Homo sapiens

<400> 11

Pro Gly Val Tyr Ala Arg Val Thr Arg Leu Arg Asp Trp Ile
1 5 10

<210> 12
<211> 24
<212> PRT
<213> Homo sapiens

<400> 12

Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Glu Glu Pro
1 5 10 15

Ser Gly Arg Phe Phe Leu Ala Gly
20

<210> 13
<211> 35
<212> PRT
<213> Homo sapiens

<400> 13

Ala Gly Tyr Leu Asp Gly Lys Val Asp Ser Cys Gln Gly Asp Ser Gly
1 5 10 15

Gly Pro Leu Val Cys Glu Glu Pro Ser Gly Arg Phe Phe Leu Ala Gly
20 25 30

Ile Val Ser
35

<210> 14
<211> 37
<212> PRT
<213> Homo sapiens

<400> 14

Ser Pro Leu Pro Phe Gly Arg His Ile Gln Pro Val Cys Leu Pro Ala
1 5 10 15

Ala Thr His Ile Phe Pro Pro Ser Lys Lys Cys Leu Ile Ser Gly Trp
20 25 30

Gly Tyr Leu Lys Glu
35

<210> 15
<211> 259
<212> PRT
<213> Homo sapiens

<400> 15

Ile Val Gly Gly Arg Glu Ala Gln Pro Gly Ser Phe Gly Ser Pro Trp
1 5 10 15

Gln Val Ser Leu Gln Val Arg Ser Gly Gly Gly Ser Arg Lys His Phe
20 25 30

Cys Gly Gly Ser Leu Ile Ser Glu Asn Trp Val Leu Thr Ala Ala His
35 40 45

```
Cys Val Ser Gly Ala Ala Ser Ala Pro Ala Ser Ser Val Arg Val Ser
    50                  55                  60

Leu Ser Val Arg Leu Gly Glu His Asn Leu Ser Leu Thr Glu Gly Thr
65                  70                  75                  80

Glu Gln Lys Phe Asp Val Lys Lys Thr Ile Ile Val His Pro Asn Tyr
                85                  90                  95

Asn Pro Asp Thr Leu Asp Asn Gly Ala Tyr Asp Asn Asp Ile Ala Leu
            100                 105                 110

Leu Lys Leu Lys Ser Pro Gly Val Thr Leu Gly Asp Thr Val Arg Pro
        115                 120                 125

Ile Cys Leu Pro Ser Ala Ser Ser Asp Leu Pro Val Gly Thr Thr Cys
    130                 135                 140

Thr Val Ser Gly Trp Gly Arg Arg Pro Thr Lys Asn Leu Gly Leu Ser
145                 150                 155                 160

Asp Thr Leu Gln Glu Val Val Val Pro Val Val Ser Arg Glu Thr Cys
                165                 170                 175

Arg Ser Ala Tyr Glu Tyr Gly Gly Thr Asp Asp Lys Val Glu Phe Val
            180                 185                 190

Thr Asp Asn Met Ile Cys Ala Gly Ala Leu Gly Gly Lys Asp Ala Cys
        195                 200                 205

Gln Gly Asp Ser Gly Gly Pro Leu Val Cys Ser Asp Gly Asn Arg Asp
    210                 215                 220

Gly Arg Trp Glu Leu Val Gly Ile Val Ser Trp Gly Ser Tyr Gly Cys
225                 230                 235                 240

Ala Arg Gly Asn Lys Pro Gly Val Tyr Thr Arg Val Ser Ser Tyr Leu
                245                 250                 255

Asp Trp Ile
```

<210> 16
<211> 43
<212> PRT
<213> Homo sapiens

<400> 16

Ser Thr Cys Gly Gly Pro Asp Glu Phe Gln Cys Gly Ser Gly Arg Arg
1 5 10 15

Cys Ile Pro Arg Ser Trp Val Cys Asp Gly Asp Pro Asp Cys Glu Asp
20 25 30

Gly Ser Asp Glu Ser Leu Glu Asn Cys Ala Ala
35 40

<210> 17
<211> 1647
<212> DNA
<213> Homo sapiens

<400> 17

atggagccca ctgtggctga cgtacacctc gtgcccagga caaccaagga agtccccgct 60
ctggatgccg cgtgctgtcg agcggccagc attggcgtgg tggccaccag ccttgtcgtc 120
ctcaccctgg gagtcctttt ggccttcctc tctacacagg gcttccacgt ggaccacacg 180
gccgagctgc ggggaatccg gtggaccagc agtttgcggc gggagacctc ggactatcac 240
cgcacgctga cgcccaccct ggaggcactg ctgcactttc tgctgcgacc cctccagacg 300
ctgagcctgg gcctggagga ggagctattg cagcgaggga tccgggcaag gctgcgggag 360
cacggcatct ccctggctgc ctatggcaca attgtgtcgg ctgagctcac agggagacat 420
aagggaccct tggcagaaag agacttcaaa tcaggccgct gtccagggaa ctccttttcc 480
tgcgggaaca gccagtgtgt gaccaaggtg aacccggagt gtgacgacca ggaggactgc 540
tccgatgggt ccgacgaggc gcactgcgag tgtggcttgc agcctgcctg gaggatggcc 600
ggcaggatcg tgggcggcat ggaagcatcc ccgggggagt ttccgtggca agccagcctt 660
cgagagaaca aggagcactt ctgtggggcc gccatcatca acgccaggtg gctggtgtct 720
gctgctcact gcttcaatga gttccaagac ccgacgaagt gggtggccta cgtgggtgcg 780
acctacctca gcggctcgga ggccagcacc gtgcgggccc aggtggtcca gatcgtcaag 840
cacccctgt acaacgcgga cacggccgac tttgacgtgg ctgtgctgga gctgaccagc 900
cctctgcctt tcggccggca catccagccc gtgtgcctcc cggctgccac acacatcttc 960
ccacccagca agaagtgcct gatctcaggc tggggctacc tcaaggagga cttcctggtc 1020
aagccagagg tgctgcagaa agccactgtg gagctgctgg accaggcact gtgtgccagc 1080
ttgtacggcc attcactcac tgacaggatg gtgtgcgctg gctacctgga cgggaaggtg 1140
gactcctgcc agggtgactc aggaggaccc ctggtctgcg aggagccctc tggccggttc 1200
tttctggctg gcatcgtgag ctggggaatc gggtgtgcgg aagcccggcg tccaggggtc 1260
tatgcccgag tcaccaggct acgtgactgg atcctggagg ccaccaccaa agccagcatg 1320
cctctggccc ccaccatggc tcctgccccc gccgccccca gcacagcctg gcccaccagt 1380

cctgagagcc ctgtggtcag cacccccacc aaatcgatgc aggccctcag taccgtgcct 1440

cttgactggg tcaccgttcc taagctacaa gtaaaaaaag aaagaaagta cgaggaactg 1500

acatatgcag gactgaaatg tgagccgtca cagccggtgg ccattaacaa tagagccaag 1560

agccccaggg acagcaactg tgggacacac ctgaactgtc tccaagcaga aatgaccagc 1620

atgcggatag ccaggagtgg ggatctg 1647

<210> 18
<211> 549
<212> PRT
<213> Homo sapiens

<400> 18

Met Glu Pro Thr Val Ala Asp Val His Leu Val Pro Arg Thr Thr Lys
1 5 10 15

Glu Val Pro Ala Leu Asp Ala Ala Cys Cys Arg Ala Ala Ser Ile Gly
20 25 30

Val Val Ala Thr Ser Leu Val Val Leu Thr Leu Gly Val Leu Leu Ala
35 40 45

Phe Leu Ser Thr Gln Gly Phe His Val Asp His Thr Ala Glu Leu Arg
50 55 60

Gly Ile Arg Trp Thr Ser Ser Leu Arg Arg Glu Thr Ser Asp Tyr His
65 70 75 80

Arg Thr Leu Thr Pro Thr Leu Glu Ala Leu Leu His Phe Leu Leu Arg
85 90 95

Pro Leu Gln Thr Leu Ser Leu Gly Leu Glu Glu Glu Leu Leu Gln Arg
100 105 110

Gly Ile Arg Ala Arg Leu Arg Glu His Gly Ile Ser Leu Ala Ala Tyr
115 120 125

Gly Thr Ile Val Ser Ala Glu Leu Thr Gly Arg His Lys Gly Pro Leu
130 135 140

Ala Glu Arg Asp Phe Lys Ser Gly Arg Cys Pro Gly Asn Ser Phe Ser
145 150 155 160

Cys Gly Asn Ser Gln Cys Val Thr Lys Val Asn Pro Glu Cys Asp Asp
165 170 175

```
Gln Glu Asp Cys Ser Asp Gly Ser Asp Glu Ala His Cys Glu Cys Gly
            180                 185                 190

Leu Gln Pro Ala Trp Arg Met Ala Gly Arg Ile Val Gly Gly Met Glu
        195                 200                 205

Ala Ser Pro Gly Glu Phe Pro Trp Gln Ala Ser Leu Arg Glu Asn Lys
    210                 215                 220

Glu His Phe Cys Gly Ala Ala Ile Ile Asn Ala Arg Trp Leu Val Ser
225                 230                 235                 240

Ala Ala His Cys Phe Asn Glu Phe Gln Asp Pro Thr Lys Trp Val Ala
                245                 250                 255

Tyr Val Gly Ala Thr Tyr Leu Ser Gly Ser Glu Ala Ser Thr Val Arg
            260                 265                 270

Ala Gln Val Val Gln Ile Val Lys His Pro Leu Tyr Asn Ala Asp Thr
        275                 280                 285

Ala Asp Phe Asp Val Ala Val Leu Glu Leu Thr Ser Pro Leu Pro Phe
    290                 295                 300

Gly Arg His Ile Gln Pro Val Cys Leu Pro Ala Ala Thr His Ile Phe
305                 310                 315                 320

Pro Pro Ser Lys Lys Cys Leu Ile Ser Gly Trp Gly Tyr Leu Lys Glu
                325                 330                 335

Asp Phe Leu Val Lys Pro Glu Val Leu Gln Lys Ala Thr Val Glu Leu
            340                 345                 350

Leu Asp Gln Ala Leu Cys Ala Ser Leu Tyr Gly His Ser Leu Thr Asp
        355                 360                 365

Arg Met Val Cys Ala Gly Tyr Leu Asp Gly Lys Val Asp Ser Cys Gln
    370                 375                 380

Gly Asp Ser Gly Gly Pro Leu Val Cys Glu Glu Pro Ser Gly Arg Phe
385                 390                 395                 400

Phe Leu Ala Gly Ile Val Ser Trp Gly Ile Gly Cys Ala Glu Ala Arg
                405                 410                 415

Arg Pro Gly Val Tyr Ala Arg Val Thr Arg Leu Arg Asp Trp Ile Leu
            420                 425                 430
```

Glu Ala Thr Thr Lys Ala Ser Met Pro Leu Ala Pro Thr Met Ala Pro
435 440 445

Ala Pro Ala Ala Pro Ser Thr Ala Trp Pro Thr Ser Pro Glu Ser Pro
450 455 460

Val Val Ser Thr Pro Thr Lys Ser Met Gln Ala Leu Ser Thr Val Pro
465 470 475 480

Leu Asp Trp Val Thr Val Pro Lys Leu Gln Val Lys Lys Glu Arg Lys
485 490 495

Tyr Glu Glu Leu Thr Tyr Ala Gly Leu Lys Cys Glu Pro Ser Gln Pro
500 505 510

Val Ala Ile Asn Asn Arg Ala Lys Ser Pro Arg Asp Ser Asn Cys Gly
515 520 525

Thr His Leu Asn Cys Leu Gln Ala Glu Met Thr Ser Met Arg Ile Ala
530 535 540

Arg Ser Gly Asp Leu
545